DOCUMENT RESUME

ED 086 260 IR 000 116

AUTHOR	Rather, John C., Ed.; Avram, Henriette D., Ed.
TITLE	National Aspects of Creating and Using MARC/RECON Records.
INSTITUTION	Library of Congress, Washington, D.C.
PUB DATE	73
NOTE	53p.
AVAILABLE FROM	Superintendent of Documents, U. S. Government Printing Office, Washington, D.C. 20402 (Stock Number 3000-00062, $2.75)
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	*Cataloging; Data Bases; *Guidelines; *Information Processing; Information Storage; *Library Automation; Library Standards; Library Technical Processes; Magnetic Tapes; Union Catalogs
IDENTIFIERS	Machine Readable Cataloging; MARC; National Union Catalog; RECON; Retrospective Conversion Working Task Force

ABSTRACT

The Retrospective Conversion (RECON) Working Task Force investigated the problems of converting retrospective catalog records to machine readable form. The major conclusions and recommendations of the Task Force cover five areas: the level of machine-readable records, conversion of other machine-readable data bases, a machine-readable National Union Catalog (NUC), an alternative strategy for RECON and general recommendations. It is concluded that: 1) full MARC (machine-readable cataloging code) format for machine-readable records is needed to distribute cataloging information, but that it is feasible to use less complete records for the NUC; 2) the projected cost of converting other machine-readable data bases are comparable to or less than those for the present per record MARC/RECON cost depending upon the method of updating; 3) data bases should be ranked by size and completeness of records; 4) standards should be established for reporting the form and contents of data bases; 5) automation of the NUC would cut time and costs and increase access points to the data. A centralized agency is recommended for large scale conversion activity of the retrospective Library of Congress records and other libraries. (JG)

FILMED FROM BEST AVAILABLE COPY

ED 086260

IR

IR 000 116

ED 086260

National Aspects of Creating and Using MARC/RECON Records

Prepared by the RECON Working Task Force

Henriette D. Avram, Chairman
Josephine S. Pulsifer
John C. Rather
Joseph A. Rosenthal
Allen B. Veaner

Edited by John C. Rather
and Henriette D. Avram

Library of Congress Washington 1973

IR 000 116

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

Library of Congress Cataloging in Publication Data

RECON Working Task Force.
National aspects of creating and using MARC/RECON records.

Includes bibliographical references.
1. MARC System—United States. I. Avram, Henriette D. II. Rather, John Carson, 1920— ed. III. Title.
Z699.4.M2R45 029.7 73-3381
ISBN 0-8444-0094-4

For sale by the Superintendent of Documents,
U.S. Government Printing Office, Washington, D.C. 20402
Price: $2.75, domestic postpaid; $2.50, GPO Bookstore
Stock Number 3000-00062

Foreword

The RECON (Retrospective Conversion) Pilot Project was initiated in August 1969 to investigate the practical problems of converting retrospective catalog records to machine-readable form. At the same time, the RECON Working Task Force began its studies of several problem areas related to the conversion of catalog records. The final report of the pilot project has been issued separately; the present publication describes the special studies. Financial support for these efforts came from the U.S. Office of Education and the Council on Library Resources, Inc. The library community has been greatly benefited by their generosity.

The rosters of the RECON Working Task Force and the RECON Advisory Committee remained essentially as they were for the RECON feasibility study; the names appear on page v. Thanks are due these persons and the institutions that allowed them to particpate. The Working Task Force wishes also to acknowledge the contributions of Barbara E. Markuson, a private consultant, who made the survey of machine-readable data bases in other libraries, and Paul E. Kebabian of the University of Vermont, who described the problems of integrating bibliographic records from various sources. Special thanks are due Susan C. Biebel of the Library of Congress for her invaluable support in all stages of these studies.

The results of these studies shed new light on several critical problems in library automation. It seems imperative that responsible persons and agencies study this report carefully and take steps to develop a national plan for conversion of retrospective catalog records that satisfies the needs of a broad community of users.

JOHN G. LORENZ,
Deputy Librarian of Congress
Chairman, RECON Advisory Committee

RECON Studies

Officer in Charge:

John G. Lorenz
Deputy Librarian of Congress

Working Task Force:

Mrs. Henriette D. Avram, *Chairman*
Library of Congress
Mrs. Josephine S. Pulsifer
Library of Congress
(formerly Washington State Library)
John C. Rather
Library of Congress
Joseph A. Rosenthal
University of California Libraries
Berkeley
Allen B. Veaner
Stanford University Libraries
(Richard De Gennaro, University of Pennsylvania Libraries, was a member of the Working Task Force for a short period but had to resign because of pressure of other responsibilities.)

Advisory Committee:

John G. Lorenz, *Chairman*
Andrew A. Aines
Senior Staff Associate
National Science Foundation
Herman H. Fussler
Professor, Graduate Library School
University of Chicago
James W. Henderson
Chief, Reference Department
New York Public Library
Frederick G. Kilgour
Director
The Ohio College Library Center
Joseph Leiter
Associate Director for Library Operations
National Library of Medicine
Maryan E. Reynolds
State Librarian
Washington State Library
Rutherford D. Rogers
Director of Libraries
Yale University
Russell Shank
Director of Libraries
Smithsonian Institution
John Sherrod
Director
National Agricultural Library
James E. Skipper
General Director
Kraus-Thomson Organization, Ltd.

Liaison Members:

Fred C. Cole
Council on Library Resources
Melvin S. Day
National Science Foundation
Burton Lamkin
Office of Education
Foster Mohrhardt
Council on Library Resources

Table of Contents

iii *Foreword*

1 *Chapter 1*: Introduction

2 *Chapter 2*: Major Conclusions and Recommendations

4 *Chapter 3*: Levels of Machine-Readable Records

7 *Chapter 4*: Conversion of Other Machine-Readable Data Bases

18 *Chapter 5*: On the Implications of a National Union Catalog in Machine-Readable Form

30 *Chapter 6*: Alternative Strategies for RECON

33 *Appendix A*: Problems in Achieving a Cooperatively Produced Machine-Readable Bibliographic Data Base

37 *Appendix B*: The National Union Catalog: Its Characteristics and Activity

42 *Appendix C*: Major Duties Involved in the Preparation of the Library of Congress Book Catalogs

44 *Appendix D*: Analysis of Library of Congress Card Orders (April 1970–March 1971)

48 *Index*

CHAPTER 1

Introduction

Concurrently with the RECON Pilot Project, the RECON Working Task Force undertook to consider certain basic questions of retrospective conversion that are of national scope.

First, is it feasible to define a level or subset of the MARC format that would allow a library using the lower level to be part of a future national network?

Second, is it possible to use machine-readable records from a variety of sources in a national bibliographic store as a way to reduce the conversion effort on the national level?

Third, what are the problems of producing a National Union Catalog from machine-readable records?

As these studies and the pilot project progressed, it also became apparent that there were many practical difficulties in carrying out a large-scale conversion project. Therefore, it seemed essential to investigate alternative strategies for RECON that might yield broad benefits in a reasonably short time span.

During the early phases of the pilot project, a task to study the problems involved in the distribution and use of name and subject cross-reference control records in machine-readable form had been outlined by the Working Task Force. This study was not initiated because of funding and timing constraints and was replaced by the study of alternative strategies.

The results of these studies are presented in the following pages. While some of the findings and recommendations are less optimistic than those of the original RECON study, it is important to realize that they still affirm the need for coordinated activity in the conversion of retrospective catalog records. Although it seems impossible to prevent all duplication of effort, it is within the realm of possibility to keep that duplication to a minimum and to achieve a high degree of compatibility among records converted in different places.

CHAPTER 2

Major Conclusions and Recommendations

The following sections give the major conclusions and recommendations of the four areas of investigation undertaken by the RECON Working Task Force.

Levels of Machine-Readable Records

Levels of machine-readable catalog records are distinguished by differences in 1) the bibliographic completeness of a record, and 2) the extent to which its contents are separately designated. The findings of this study were:

1) The level of a record must be adequate for the purposes it will serve.

2) In terms of national use, a machine-readable record may function as a means of distributing cataloging information and as a means of reporting holdings to a national union catalog (NUC).

3) To satisfy the needs of diverse installations and applications, records for general distribution should be in the full MARC format.

4) Records that satisfy the NUC function are not necessarily identical with those that satisfy the distribution function.

5) It is feasible to define the characteristics of a machine-readable NUC report at a lower level than the full MARC format.

Conversion of Other Machine-Readable Data Bases

Machine-readable bibliographic data bases do exist that could be used to increase the volume of the national store under the following conditions:

1) The per-record cost of converting these records to the MARC format, comparing them with records in the LC Official Catalog, and updating their content to the point where they match those records approaches the present per-record MARC/RECON cost.

2) The cost of converting the same records if only the access points were updated appears to be substantially lower than present MARC/RECON costs. The minimum cost of this method of data base conversion is probably on the order of one-half of present costs. Since these data could not be used in this form by the Library of Congress, the question of how this effort could be funded remains to be resolved.

3) Should any such program be undertaken, the high potential data bases should be ranked by size and completeness of content of records. However, the character of the records would have to be evaluated to determine whether the estimated per-record conversion cost held true for any given data base.

4) A standard should be established for reporting the form of material, language, and the content of machine-readable records in library data bases to simplify the job of determining the utility of another library's data base.

A National Union Catalog in Machine-Readable Form

Automation of the National Union Catalog using the register/index form would have the following advantages:

1) The range of access points to the bibliographic data would be extended to titles and series.

2) All types of indexes would be cumulated and published on the same schedule.

3) The time required to produce cumulations would be significantly reduced.

4) The cost of the automated system offering these advantages for monthly, quarterly, and annual issues would not exceed the cost of the present manual system. The cost of producing the quinquennial would be sharply reduced.

5) The cost of the automated system would gradually be reduced as more languages are covered by the MARC Distribution Service. Further cost reductions may be possible as other libraries are able to report their holdings in machine-readable form.

6) Converting NUC reports and master index records for LC non-MARC records to machine-readable form would create a data base that could be searched by nonconventional access points (e.g., language, imprint date, geographic area).

7) The NUC data base might eventually form the nucleus of an on-line network of regional bibliographic centers.

An Alternative Strategy for RECON

There is no ideal strategy for large-scale conversion of retrospective catalog records. The critical questions of the languages to be covered, the dates of the records, the forms of material, the extent of bibliographic information, and the details of the machine format yield widely different answers depending on the type and size of the library involved. Therefore:

1) A centralized agency or component of an agency should be established expressly to undertake a large-scale conversion activity. This effort should not divert the Library of Congress from its present objective of going forward as rapidly as possible to convert all of its current catalog records to machine-readable form. To the extent that retrospective records are required for Library of Congress purposes (e.g., Card Division mechanization; special book catalogs), LC would convert these records according to its present practices.

2) The central agency should have two major functions:

a. It should undertake a program to convert the retrospective LC records that are most in demand. Initially, the criterion for selection might be those records ordered from the LC Card Division more than a specified number of times.

b. It should be responsible for adapting machine-readable records from libraries other than LC. The scope of this cooperative approach would be modified as each new language is covered at LC.

In developing its program and carrying out these tasks, the agency should draw on the experience gained in the MARC and RECON activities at the Library of Congress. Since users will be obtaining current catalogs from the Library of Congress, it is essential that the products of these two enterprises be entirely compatible.

3) To ensure that the conversion of other libraries' machine-readable data bases results in consistent records, the following procedures are recommended:

a. If a library converts, it should use the best available LC record.

b. If at all possible, the full MARC format should be used.

c. The centralized agency should undertake to process records to bring them to the full MARC format (if necessary) and to make the access points compatible with the LC Official Catalog.

General Recommendation

The problem of conversion of retrospective records to machine-readable form is of concern to all types of libraries in all parts of the country. Therefore, the National Commission for Libraries and Information Science should review the present report as well as the original RECON feasibility study to determine the course of action that is in the national interest. The Commission might also explore the sources of funds to implement its recommendations for a national program for the conversion of retrospective catalog records.

CHAPTER 3

Levels of Machine-Readable Records

This study reports the conclusions reached with respect to the feasibility of determining a level or subset of the established MARC content designators that would still allow a library using it to be part of a future national network.

Definition of "Level"

During the initial RECON study the Working Task Force, for discussion purposes, considered levels of encoding detail of machine-readable catalog records in relation to the conditions under which conversion might occur. A level was distinguished by differences in 1) the bibliographic completeness of a record, and 2) the extent to which its contents were separately designated. With respect to the latter point, the RECON report stated:

> A machine format for recording of bibliographic data and the identification of these data for machine manipulation is composed of a basic structure (physical representation), content designators (tags, delimiters, subfield codes), and contents (data elements in fixed and variable fields). Although the basic structure should remain constant, the contents and their designation are subject to variation. For example, a name entry could be designated merely as a name instead of being distinguished as a personal name or corporate name. When a distinction is made, a personal name entry can be further refined as a single surname, multiple surname, or forename. Likewise, if a personal name entry contains date of birth and/or death, relationship to the work (editor, compiler, etc.), or title, these data elements can be identified or can be treated as part of the name entry without any unique identification. Thus individual data elements can be identified at various levels of completeness.[1]

Appendix F of the RECON report tentatively defined three levels:

> *Level 1* involves the encoding of bibliographic items according to the practices followed at the Library of Congress for currently cataloged items, i.e, the MARC II format. A distinguishing feature of level 1 is the inclusion of certain content designators and data elements which, in some instances, can be specified only with the physical item in hand.
>
> *Level 2* supplies the same degree of detail as in level 1 insofar as it can be ascertained through an already supplied bibliographic record. . . .
>
> *Level 3* would be distinguished by the fact that only part of the bibliographic data in the original catalog record would be transcribed. In addition, content designators might be restricted . . .[2]

At the outset of the present study, however, it was recognized that incomplete bibliographic description is not acceptable in records for national use. In addition, it seemed that the question of having a level below level 2 really arose from a desire to define a machine-readable record with a lesser degree of content designation rather than one with less bibliographic data. It was decided, therefore, to concentrate the study effort on this task, and the original formulation of level 3 was discarded.

On further consideration, it was realized also that the distinguishing feature between levels 1 and 2 was not significant. Omission of data elements that cannot be determined unless the book is in hand may simplify an individual record but does not simplify the content designators in the format because these elements are often present in other records. Thus, as far as content designation is concerned, levels 1 and 2 (as originally defined) were in fact the same.

Once this similarity became apparent, it was recognized that the specification of levels really depended on the functions of machine-readable catalog records from the standpoint of national use.

Functions and Levels

On the basis of present knowledge, it seems that machine-readable records will serve two primary

functions for national use. The first involves the distribution of cataloging information in machine-readable form for use by library networks, library systems, and individual libraries; the second involves the recording of bibliographic data in a national union catalog to reflect the holdings of libraries in the United States and Canada. In this report, the first is called the distribution function; the second is called the national union catalog (NUC) function. Each of these functions can be related to a distinct level of machine-readable record.

The Distribution Function

The distribution function can best be satisfied by a detailed record in a communications format from which an individual library can extract the subset of data useful in its application. At the present stage of library automation, it is impossible to define rigorously all of the potential uses of machine-readable catalog records. Thus, there is no way to predict which data elements may not be needed or to rank them according to their value to a wide variety of users under different circumstances.

To confirm wide variation in treatment of the MARC formats, an analysis was made of the use of MARC content designators by eight library systems and emerging networks. The data from this analysis were synthesized for presentation in two tables. Table 3.1 shows the acceptance of content designators in terms of the absolute number of libraries using them. It should be read as shown by the following examples: 1) 26 of the 63 MARC tags are used by all eight libraries; 2) 92 of the 126 indicators are used by three libraries. Table 3.2 shows the acceptance of content designators in relative terms. Thus, if only three libraries were using a particular tag and all used the associated subfield codes, the acceptance of those subfield codes was calculated as 100 percent. In both Tables 3.1 and 3.2, the columns on indicators and subfield codes include responses only from those libraries that were definitely using the tag with which a given indicator or subfield code was associated. The analysis excludes tags for which no immediate implementation is planned by the MARC Distribution Service.

The major findings of this analysis may be summarized as follows:

1) Of 19 fixed fields, 14 were used by at least half of the libraries and all were used by at least one library.

TABLE 3.1—*Use of MARC content designators by 8 library systems or networks*

Number of libraries	Number of items			
	Fixed fields [1]	Tags	Indicators [2]	Subfield codes [2]
Total	19	[3] 63	126	181
8		26		1
7		6		88
6		3	2	45
5	1	5	7	15
4	6	3	9	9
3	7	2	92	11
2	4	4	16	9
1	1	7		3
None		7		

[1] Only 6 libraries supplied this information.
[2] This column includes responses only from those libraries that were definitely using the tag with which a given indicator or subfield code was associated.
[3] Excludes tags for which no immediate implementation is planned.

TABLE 3.2—*Percentage of acceptance of MARC content designators by 8 library systems or networks*

Percent of libraries	Number of items			
	Fixed fields	Tags	Indicators [1]	Subfield codes [1]
Total	19	[2] 63	126	181
100		26		10
75 to 99	1	9	2	134
50 to 74	13	8	16	32
25 to 49	4	6	108	5
1 to 24	1	7		
0		7		

[1] This column includes responses only from those libraries that were definitely using the tag with which a given indicator or subfield code was associated.
[2] Excludes tags for which no immediate implementation is planned.

2) Of 63 tags, 43 were used by at least half of the libraries and 26 were used by all of them. Seven tags were not used by any of the libraries studied, but these tags cover items that will appear in machine records produced by the National Library of Medicine, the National Agricultural Library, and the British National Bibliography.

3) Of 126 indicators, only 18 were used by at least half of the libraries. The highest degree of acceptance was the use of the same two indicators by six libraries. On the other hand, each indicator was used by at least two libraries.

4) Of 181 subfield codes, 176 were used by at least half of the libraries that were using the related tags. Each subfield code was used by at least a quarter of the libraries that could express a relevant opinion.

The foregoing analysis confirmed the view that a nationally distributed record should be as rich in content designation as possible. Failure to provide this detail would result in many libraries having to enrich the record to satisfy local needs, a process more costly than deleting items selectively. Therefore, as of now, the present MARC format constitutes the level required to satisfy the national distribution function.

The National Union Catalog Function

As noted above, the NUC function relates to the use of machine-readable records to build a national union catalog. At first thought, it might appear that this function overlaps the distribution function. As far as Library of Congress cataloging is concerned, this view is correct. It is valid also with respect to cooperative cataloging entries issued by the Library as part of the card service. However, the two functions are quite distinct as far as regular reports to NUC are concerned.

The essential difference between the two categories of catalog records is that those issued as LC cards have been completely checked against the Library's authority files and edited for consistency, whereas only the main and added entries of NUC reports have been checked for compatibility. The impact of this difference can be judged from the fact that an attempt to distribute NUC reports as proof slips several years ago was abandoned because the response to this service did not justify its continuance.

Distributing NUC reports in machine-readable form would add another dimension to the problem of processing them, because, to be flexible enough for wide acceptance, NUC reports would have to be entirely compatible with those issued by the MARC Distribution Service. Since compatibility would involve more detailed content designation than many libraries might put into their records for local use, libraries would have to be willing to provide this detail in NUC reports, or the level of NUC reports would have to be upgraded centrally. As the certification of the bibliographic data and the content designators would entail a major workload for the Library of Congress, it does not seem practical to pursue this goal at present.

It is possible, however, to define a subset of content designators to cover the eventuality that outside libraries may be able to report their holdings to NUC in machine-readable form. A MARC subset can be determined for the NUC function because this function involves processing records in a multiplicity of places to be used centrally for specifically definable purposes. The distribution function, on the other hand, involves the preparation of records at a central source to be used for a wide variety of purposes in a multiplicity of places. The difference is vital when it comes to stating the requirements for the two types of records.

The specifications of a machine-readable record to fulfill the NUC function depend on the nature and functions of the national union catalog itself. The content designators for such a record were defined in a separate investigation which is described in Chapter 5. The present study was considered to be completed once the feasibility of defining a level of machine-readable record for that purpose was established.

Conclusions

The findings of this study of the feasibility of defining levels of machine-readable bibliographic records are as follows:

1) The level of a record must be adequate for the purposes it will serve.

2) In terms of national use, a machine-readable record may function as a means of distributing cataloging information and as a means of reporting holdings to a national union catalog.

3) To satisfy the needs of diverse installations and applications, records for general distribution should be in the full MARC format.

4) Records that satisfy the NUC function are not necessarily identical with those that satisfy the distribution function.

5) It is feasible to define the characteristics of a machine-readable NUC report at a lower level than the full MARC format.

References

[1] RECON Working Task Force. Conversion of retrospective catalog records to machine-readable form. Washington, Library of Congress, 1969, p. 43.

[2] Ibid., p. 164.

CHAPTER 4

Conversion of Other Machine-Readable Data Bases

Introduction

A large pool of machine-readable records has accumulated as a result of automation projects in various libraries. There is widespread opinion that these records could be used to build a national bibliographic data base. The potential benefits are thought to be avoidance of duplication of input, more rapid creation of a large store, and reduction of the manpower required to accomplish this task with a consequent lowering of the cost. Presumably many of the records now in machine-readable form were derived from MARC records distributed over the past several years. However, since MARC has covered only recent English language materials, this pool of machine-readable bibliographic records must include a large number of titles not currently available in MARC.

Counterbalancing the possible advantage of using these records is the fact that they are known to vary considerably in terms of their bibliographic content and their machine format. Thus, they would have to be processed centrally to allow them to be integrated with records being produced by the Library of Congress. To determine what problems would be encountered in this processing, the RECON Working Task Force undertook to gather information about representative machine-readable data bases and to assess their potential for this purpose. The task was divided into two phases. Phase I was a survey of existing library data bases and an analysis of the machine-readable records as to bibliographic content and format compatability with MARC. Selected data bases became candidates for further analysis. Phase II was the analysis of the costs and methods (when applicable) to utilize data bases from various sources (selected data bases from Phase I) to build a national bibliographic store and a comparison of these costs with the costs of RECON conversion at the Library of Congress.

Phase I—Survey of Existing Library Data Bases

Methodology

A list of libraries having cataloging data in machine-readable form was compiled. Some of these were already known to exist; some were identified by a review of the library literature. The report entitled *Book Form Catalogs: A Listing*,[1] prepared by the ALA RTSD Book Catalogs Committee, was useful in identifying some of the less well-known data bases.

The following criteria were established for selection of data bases to be surveyed:

1) The data base had to include records for monographs.

2) Data bases known to have fewer than 15,000 records were excluded.

3) Data bases known to be entirely or predominantly based on MARC Pilot Project or MARC records were excluded.

4) Data bases had to be potentially available to RECON. This eliminated data bases with security restrictions and most commercial data bases.

No attempt was made to be exhaustive in identifying existing data bases meeting these criteria. The purpose of the study was to investigate the overall problem. If use of outside data bases is judged feasible, a more comprehensive survey can be undertaken. Thus, failure to consider a particular data base does not necessarily mean that it might not meet the above criteria and be potentially useful to RECON. However, the RECON Working Task Force is reasonably sure that most data bases meeting these criteria were examined.

A two-step survey was undertaken. The first survey elicited information that would serve to discriminate data bases of low potential utility from data bases that warranted further study. A general questionnaire requested information on availability, size and composition, the data elements in the catalog record format, the character set used, and the source of the cataloging data upon which the machine-readable input was based. The questionnaire was sent to 42 libraries, 33 of which responded.

When the questionnaire returns were analyzed, four data bases were judged to be outside the scope of the study and seven others were excluded from further consideration because some were quite small and others contained only brief catalog records. Although other factors were considered, the 22 libraries selected for the follow-up survey were chosen primarily on the size of the data base and the fullness of the catalog record. These libraries were asked to submit additional information including format documentation, sample catalog cards, sample input worksheets, etc. The information requested was, in general, supplied from two different sources. Bibliographical materials were supplied by a bibliographical resource person who had been specified by the respondent on the initial questionnaire; similarly, technical data were supplied by the designated technical resource respondent.

A worksheet was prepared to reduce all the documentation provided for each system to a standardized form. This worksheet provided for a generalized description of data base characteristics based on the initial questionnaire response, a brief summary of the major features of the format, a field-by-field comparison of the local record with the MARC format, and a sample catalog output.

Analysis of Machine-Readable Formats

The evaluation of each format in terms of potential usefulness was made on the basis of submitted documentation and, in some cases, by follow-up telephone inquiries. Since many of the formats were relatively complicated and some ambiguities existed, errors in interpretation may have been made. It is believed, however, that changes in minor details would not affect the major findings of this study.

Analysis and comparison of 22 machine-readable catalog formats is a sobering experience. The variation among them was greater than had been anticipated. In the beginning, it was assumed that some basic patterns would be discovered and that these would provide the overall framework for the analysis. Attempts to discover these basic relationships were fruitless, however. In the end, although the format of each data base was compared to MARC, the data bases were categorized more from the point of view of bibliographical completeness than according to technical characteristics.

Analysis was made difficult also by the nature of the documentation, the imprecision of terminology, and the lack of clear data field definitions. Each of these points is briefly discussed below.

The amount of documentation supplied for the data bases ranged from extremely detailed to extremely sparse. In most cases the available documentation consisted of bits and pieces, but some libraries provided well-organized, logical, and unified documentation. Generally, the lack of sufficient documentation was a serious handicap to detailed format analysis.

Both technical and bibliographical terminology presented problems. Terms such as "title paragraph" were frequently used, but the scope of the terms often differed widely. For example, sometimes the title paragraph included the edition statement and sometimes the latter was considered as a separate field. Fields named "topical subjects," "subject headings," and "subject tracings" had to be examined against the input records to try to determine how they were defined for a particular format. In only a few instances were the data fields clearly defined with examples, scope, and limits explicitly stated. Nonstandard and obviously local terminology also presented problems.

In general, the format descriptions were more detailed with respect to those data fields associated with control and housekeeping information than they were for bibliographical information. In some instances the system documentation merely indicated the bibliographical portion of the format as "variable field data." In such cases, the variable fields had to be determined from an examination of the sample input and output documents and the responses on the questionnaire. The danger of this is that the samples supplied provide only a limited number of records and probably do not illustrate the possible range of fields included.

It would be theoretically possible to rank formats on a weighted basis from "most like MARC" to "least like," but the analysis would be extremely complex and costly because of the large number of variables involved. The recommendations in this

study are based on a subjective ranking of the formats. In making the recommendations an attempt was made to keep the following variables in mind for each format: completeness of the bibliographical data, structure of the format, size of data base, nature of the library, future growth potential, proportion of non-MARC records, character sets used, and nature of catalog source (e.g., local, Library of Congress, commercial vendor).

For each format, the data fields were compared to the MARC format on the basis of the following conditions: 1) field present in both formats, 2) field not present in the local format and not capable of being generated from other data in the record, and 3) field not present in local format and judged to be capable of being generated from other tagged data in the record or by use of a format recognition algorithm. The evaluation did not include a fourth condition: data fields present in the local record and not provided for in the MARC format. In most cases, these local fields were tagged and it was assumed that a conversion program could strip these elements automatically.

The data bases were divided into three groups after the analysis was completed: high potential, medium potential, low potential. It must be emphasized that these value judgments are made only with respect to RECON needs and are in no way meant to reflect on either the quality of the local collection, system, or data base or on the suitability of a particular format for a given library's needs. Although it is difficult to define rigidly the differences between the three groupings, the major characteristics of the data bases in each group are summarized below.

High-Potential Data Bases

Although this group does not comprise the largest aggregate of records, it should become the largest source of unique titles when some planned conversion projects are completed.

The records are similar to LC MARC records in terms of fullness of catalog entry, record structure, and discrimination of fields and subfields. Since most of the formats for these data bases were developed during or since the MARC projects, they incorporate many MARC-like features, e.g., a fixed field segment, a record directory, and a variable field segment. With one exception, the records contain upper and lowercase characters and most of the character sets include diacritical marks.

Medium-Potential Data Bases

With one exception, the data bases in this group contain fewer than 100,000 records and average about 50,000 items each. They tend to have fairly full catalog entries, although bibliographic notes, illustration statements, and size indication are usually excluded. Most of the entries are based on LC cataloging, edited to conform to local practices, and most have LC subject headings and many have LC classification numbers. Some of the data bases compare favorably with the top group in terms of bibliographical completeness but a few omit certain fields, such as place of publication, series notes, and bibliographical notes. Five of the data bases include the LC card number as part of the record although this field is not present in every record.

The distinction between these data bases and the high-potential group is primarily attributable to variations from the MARC format and the absence of a rich tagging and coding of the data. The formats tend to be more sophisticated than those in the low-potential category and generally the format provides for some fixed field codes in addition to the standard cataloging data. Three of the libraries have data bases encoded in uppercase character sets, but the majority use upper and lowercase characters.

Low-Potential Data Bases

These data bases cannot be characterized by size: they range from about 15,000 entries to 500,000, and, therefore, include some of the largest data bases for monographs in existence. They can be characterized in terms of fullness of monograph entry. Most of them include only the main entry, title (sometimes only a short title), brief imprint, and local call number. Even those data bases which contain fuller entries usually eliminate details such as bibliographical notes, illustration statements, size, and series notes and tracings.

The formats used by libraries in this category are generally quite simplified. Almost no fixed field codes describing the cataloged item are included since the record is usually limited to those data fields required for brief entry book catalogs and for circulation purposes. Most of these data bases are encoded in uppercase character sets.

Most of the data bases in this group were eliminated on the basis of the initial questionnaire return; a few were eliminated after more detailed information was supplied.

479-312 O—73——2

Findings

The aggregate data base of the 29 in-scope respondents amounts to more than 3,720,000 records of all types, including about 2,500,000 records for monographs. Of course, these figures do not represent unique records but it was impossible to estimate the amount of duplication among data bases. The cost of producing this store of records is difficult to estimate precisely as a number of different methods were used by a variety of organizations—many of which were of necessity investing a substantial part of their efforts in learning the required techniques of the field. It seems unlikely, however, that the average cost is less than $1.00 per record. Thus, an investment on the part of the library community of several million dollars has been expended. These are impressive totals, especially when one considers that some libraries did not respond to the questionnaire.

Table 4.1 shows the aggregate number of records in each category of data base and the number of new records added per year. Annual additions to the data base were reported only by those 22 libraries taking part in the intensive survey.

These figures are evidence of the tremendous expenditure of manpower and money that has already gone into the creation of machine-readable data bases. A small segment of the library community has been able to create a substantial machine-readable data base within a few years.

From the standpoint of standardized bibliographic control, the picture is less favorable. The bibliographic variations among records can be readily seen in Table 4.1. In general, the high potential group shows the greatest bibliographical conformity, but even here there are significant differences as well as many differences in format structure and tagging.

It appears that the more recently a data base was created, the more bibliographically complete it is and the more flexible the format tends to be. Thus, the influence of the MARC format is beginning to be felt. Of the 29 data bases reported, 22 use non-MARC formats, 2 use the MARC Pilot Project format, and 5 use a format based on or identical to the MARC format or are planning to convert to this format. A few of the non-MARC formats were (in the opinion of the respondents) compatible or convertible to MARC, but no respondent reported any actual attempt at such a conversion.

No data base was discovered that was identical to the LC MARC data base from a technical viewpoint although some are nearly so. Most data bases depart from standard LC catalog information by modifying some data fields locally and adding or eliminating others.

Two high potential data bases with a high degree of compatibility with MARC and a medium potential data base that differs from MARC both in the level of content designators and in bibliographical completeness were selected for the Phase II analysis.

TABLE 4.1—*Characteristics of 29 machine-readable bibliographic data bases, by RECON use potential*

Characteristics	High potential (11)	Medium potential (8)	Low potential (10)
Size	636, 000	587, 000	1, 302, 000
Number of items added per year [1]	333, 000	78, 000	[2]
Source:			
MARC	3	1	0
LC full cataloging	7	4	0
LC cataloging modified	6	4	5
Local-full cataloging	10	1	0
Local-brief cataloging	0	4	7
Format:			
MARC Pilot Project	2	0	0
MARC	5	0	0
Non-MARC	4	8	10
Language:			
English	11	8	10
Foreign	11	6	5
Data elements:			
Main entry	11	8	10
Title	11	7	3
Short title	11	5	8
Edition	11	8	6
Imprint-full	10	3	0
Imprint-brief	3	4	9
Collation	11	5	3
Series notes	11	6	2
Other notes	11	7	0
LC subject	9	8	5
Non LC subject	5	1	0
Added entry	11	7	3
Decimal classification	1	4	6
LC classification	9	5	1
Card number	10	5	0
Character set:			
Uppercase only	1	3	7
Upper and lowercase	10	5	3
Diacritics	7	3	1

[1] Based on 22 responses.
[2] Incomplete.

SURVEY PARTICIPANTS

Abel, Richard & Co., Inc.
Air Force Cambridge Research Laboratory Library
Austin (Texas) Public Library
Baltimore County (Maryland) Public Library
Black-Gold Cooperative Library System, Ventura, California
Chester County (Pa.) Library System
Cuyahoga (Ohio) Community College Library
Eastern New Mexico University
Georgia Institute of Technology Library
Harvard University, Widener Library
Honnold Library for the Claremont Colleges, Claremont, California
Indiana University of Pennsylvania, Indiana, Pa.
Jefferson County (Colorado) Public Library
Montgomery County (Maryland) Library
National Agricultural Library
National Library of Medicine
New York Public Library (Branch, Dance, & Research)
Redstone Scientific Information Center
San Antonio College
Stanford University Undergraduate Library
SUNY Upstate Medical Center Library
Tennessee State Library
University of California, Santa Cruz
University of California, Union Catalog, ILR
University of Chicago Library
University of Vermont, Dana Medical Library
Vancouver Island (British Columbia) Regional Library
Washington State Library
Yale University Medical Library

Phase II—Analysis of Costs and Methods

The attraction of using machine-readable records from other libraries as input to the LC MARC data base lies largely in the fact that such a procedure would eliminate at least the need for original keyboarding of the record. Against this obvious saving in labor, one has to measure the costs of acquiring the data base from the originating library, converting the format to MARC, searching the LC Official Catalog, and updating the records. The purpose of this phase of the study is to analyze what is required to convert the data bases of two representative libraries selected as the result of Phase I and to estimate the cost of doing so.

Programming Costs

A fixed cost of using each non-MARC data base for the purpose of adding records to the LC MARC data base is the cost of programming to convert the record format of the given data base to LC MARC format. This cost is not strictly fixed as the larger the data base, the more worthwhile it might be to add certain features to the program to take care of special cases that for smaller data bases might be more economically done manually. In addition, the experience gained as each program is written should tend to reduce the effort and, therefore, the cost of subsequent programming. However, in this analysis, the primary cost of programming format modifications is considered to be independent of data base size and experience.

There are two methods of converting the records of an outside library to the MARC format:

1) The data can be converted directly to the MARC format.
2) The data can be converted to the input format required for the LC format recognition programs (data strings without content designators).

It would appear that the choice of method would depend on the degree of similarity between the input data base and the MARC format. However, even with a close approximation to the MARC format, it might pay to use format recognition whenever the bibliographic data (including punctuation) are taken from LC catalog cards because the sophistication of the format recognition programs enables them to perform with remarkable accuracy. Since much of the same developmental work might be necessary to write a program to convert an input tape in another format to MARC, such a program would be costly to develop. Although a program to convert a record to the input format required for format recognition may be complicated by the tagging scheme of the other library's records, the fact that the format recognition programs are operational offers the possibility of a great saving in programming time. It should be emphasized that the success of such a program is dependent on the degree of explicit identification of the data elements in the input record; that is, the extent to which they resemble MARC content designators.

In the final analysis, each data base must be studied individually before a method of conversion can be chosen. Since the problems associated with either approach have similar characteristics, it was assumed that the technique would be conversion to the input format required for format recognition.

Published information on the cost of designing, writing, and testing computer programs is surprisingly sparse. Dolby [2] reports that a primary factor in estimating programming cost is the size of the program because the cost tends to increase as the square of the size of the program rather than as a linear function. Theoretical support for such an argument can be made by observing that the number of possible interactions between pairs of instructions (and hence, possible program errors) increases as the length of a program increases. Thus, the well-known advantage of modularizing programs through the use of subroutines, macros, etc., can be explained by the fact that modularity reduces a large program to a sequence of small subprograms, which has the effect of reducing the number of interactions among instructions.

Aron mentions four techniques in his article "Estimating Resources for Large Programming Systems." [3] Two of these techniques, the Constraint Method and the Units of Work Method, are not applicable because they vary or subdivide the task to fit the available manpower. The other two, the Quantitative Method and the Experience Method, are worth considering for this study.

The approach used in the Quantitative Method is to estimate the size of a desired program (instructions or lines of code) by comparing program requirements with those of similar projects. For a large project, individual estimates for sub-unit at various levels can be used. The total number of instructions is subdivided into three classes: easy, medium, and difficult. The number of man-hours of programming required is computed per man day at the rate of 20 instructions for easy programming, 10 instructions for medium programming, 5 instructions for difficult programming. The hours of direct labor obtained this way are adjusted to allow for supervision and other overhead factors and converted into costs by applying appropriate rates.

The approach used in the Experience Method is to estimate the cost, size, and time requirements of a programming project by comparing it with similar previous ones. Although this method is inexact, it is widely used because of its practicality. Size and man-hour figures are available for a program that has been written to convert MARC records to an input format to test format recognition processing. This program converts a MARC record to a record containing data strings without content designators. When this record is processed through format recognition, the result should be a record identical to the one originally converted to data strings. The program was written in Assembly Language Coding; it contains 1,520 lines of code. It took 555 man-hours of programming. Assuming a cost of $18.00 per hour for contractual programming, the program cost was approximately $10,000. While it is true that the number of lines of coding and ultimately the cost of the program depend on the programming language used and the competence of the programmer, it was felt that the direct experience with an almost identical problem justified using the LC estimate.

Processing Strategy

The steps by which the conversion of another library's data base would proceed depend on the end in view. Two objectives are possible:

1) To obtain a record in which the access points would be identical with those on the record in the LC Official Catalog but the other data would be as furnished by the library that created the record. This approach assumes that the record is essentially bibliographically complete.

2) To obtain a record that is identical with one in the LC Official Catalog; such a record would be the equivalent of one supplied by the LC MARC Distribution Service.

In devising methods to achieve these objectives, it was assumed that records which have no match in the LC Official Catalog would not be added to the file. Attempting to edit non-LC records for inclusion would add a major cost as is shown by the study of the requirements for a national union catalog in machine-readable form (see Chapter 5). The decision to disregard nonmatching records in the present study was made on practical grounds: there is no convenient way to estimate their proportion in any given data base.

For the method to achieve Objective 1, the following assumptions were made:

1) Main, added, and subject entries would be changed as necessary to make them identical with the corresponding LC headings used for the same record.

2) Data on an LC card that are lacking in the other library's record would not be added, except for LC card number, international standard book number (ISBN), and the LC call number.

3) If the other library's record has access points

not on the LC card, they would be excluded.

4) Proofing would amount to 50 percent of the cost of proofing regular MARC/RECON records because proofing will be primarily to detect format recognition errors and to confirm the accuracy of the data in the access points.

5) The cost of correction typing would be equal to that of making corrections on MARC/RECON records because it is assumed that catalog comparison of access points will not result in any more changes than are now made on RECON records and that other types of errors (e.g., typographical errors, format recognition errors) would be at the same level.

6) The cost of verification (the final reading for content) would be 50 percent of the cost of verifying MARC/RECON records because only the accuracy of the content designators and the primary access points would have to be verified.

For the method to achieve Objective 2, the following assumptions were made:

1) All data on LC records that was lacking in the other libraries would be added.

2) Data elements not on the LC card would be deleted; data elements that differ would be changed to match the LC card.

3) Although the basic proofing cost would be the same as the present MARC/RECON cost, ensuring that the entire record matched the LC record in every detail would involve extra expense. No attempt was made to estimate this cost because it would depend on whether the proofing was done at the Official Catalog or at a later stage from a copy of the LC record. It is probable, however, that the cost of performing this task would offset the saving of the cost of original keying.

4) The workload (and, therefore, the cost) of typing corrections would be twice that of Method no. 1.

5) The cost of verification would be the same as the present MARC/RECON cost.

It will be observed that the costs in this method are essentially those of MARC/RECON; only the cost of original keying would be saved. This is because certification of an outside library record as an LC record requires the complete proofing and verification process and the proportion of corrections (i.e., fields added or changed) would almost certainly be greater than is true for a record converted directly at the Library of Congress.

As noted above, in processing an outside library data base, it is necessary to eliminate all records already in the MARC data base as well as those outside its present scope. The procedures for doing this would be heavily dependent upon the nature of the data base being processed.

For the purpose of this analysis, records eligible for selection from a data base may be defined as records represented in the LC Official Catalog, but not yet included in the MARC data base. Ineligible records include those that duplicate existing MARC records, records not represented in the LC Official Catalog, and records for forms of material not yet included in the MARC Distribution Service. Language would not be grounds for declaring a record ineligible. Identification of ineligible records should be done by computer when feasible but, in many cases, eligibility can be determined only by manually checking the records against the LC Official Catalog.

The cost of machine searching varies with the amount of manipulation of fields required to derive the search argument. Manual searching against the Official Catalog costs approximately $.10 per record. Since the rental cost of the LC IBM 360/40 configuration is $27,767 per month or $.0438 per second, based on 176 hours per month, the manual searching cost is approximately equal to the cost of 2.3 seconds of machine time. This exceeds the time required even for a relatively complex machine search. Therefore, machine procedures should be used, wherever possible, to decrease the number of manual searches required against the Official Catalog.

The following statements suggest various techniques that might be used. It may be expedient to process the source data against the MARC data (matching on LC card number, if available, or an author/title search code) to eliminate records already in the MARC data base. Again, depending on the characteristics of the source data base, automatic algorithmic deletion by language and imprint date (e.g., English language records with publication date 1968 or later) might be an alternative technique to searching to eliminate records already in MARC. Likewise, if the source data came from a library that used LC cataloging data when available and always included the LC card number

in machine-readable records, it might prove expedient to automatically delete records without LC card numbers based on the assumption that the lack of an LC card number in the source record would be good evidence that the record would not appear in the Official Catalog. This technique would reduce manual searching with only slight risk of losing records that should have been included. Naturally, the validity of this assumption would have to be tested in any given situation.

It must be kept in mind that the expected yield of ineligible records as the result of a machine search has an important bearing on whether it should be made. For example, comparison on LC card number against the MARC data base would not be economic for a large retrospective (i.e., before 1968) data base even if LC card numbers were given, because the disk lookup against a table of LC card numbers for each record in the file would yield few records for a large cost. In this situation, algorithmic deletion by language and publication date might be more appropriate. Assuming that the characteristics of the data bases are known by their creators, the need for an individual determination of the particular strategy for eliminating records for that data base should not pose a major problem.

Depending on the characteristics of the data base, the automatic deletion of ineligible records might precede or follow the conversion of the records to the input format required for format recognition. Records would be processed through format recognition, printed, searched to delete records that are not eliminated automatically, and (when eligible) compared manually against the matching LC records. After proofing for the accuracy of format recognition, the records would be modified to reflect corrections of content designators and any data changed as a result of catalog comparison. The records would be printed again for proofing and final verification prior to being added to the MARC data base.

Procedure Costs

Computer run costs have not been included in the cost estimates. Analysis indicates that the per record cost of original conversion at LC (input format from the DigiData converter to input for format recognition) is approximately equal to the cost of converting a record from another data base to the input format for format recognition. Since these costs are not reported as part of the RECON cost estimates, they have not been included as part of the present estimates.

It is recognized that some records processed in this phase of the conversion may be discovered to be ineligible at a later stage. Strictly speaking, the cost of processing these ineligible records should be prorated among the eligible records. This has not been done, however, because there is no valid way to estimate the percentage of ineligible records discovered this way and, in any case, the incremental cost would probably be insignificant in relation to the total average conversion cost.

Programming cost is a function of the number of usable records and must be apportioned accordingly. The cost of programming to convert records of a large research library can be expected to be amortized over an indefinite period since such a data base will continue to yield records of value to a national data base. On the other hand, the cost of programming to convert records of a smaller library may warrant apportionment only over the number of records in a one-time conversion effort, if that library's future acquisitions are unlikely to contribute significantly to the national data base.

Representative Conversion Costs

The generalized conversion strategies were applied to two high potential data bases identified in the survey described earlier in this chapter; they were the University of Chicago Library and the Research Libraries of the New York Public Library. When the medium potential data base was examined, it was found that the records were not bibliographically complete enough to be suitable for a national data base without adding or changing many data elements. Thus, Objective 1 would not be appropriate for this data base and Objective 2 would entail costly updating. Therefore, it was decided to limit the cost analysis to the two high potential data bases.

Table 4.2 shows the basic characteristics of these two data bases. Since both of them contain records that are substantially like those in the LC Official Catalog, it is feasible to consider both of the conversion objectives that have been described earlier. The basic steps for the University of Chicago Library data base would be as follows:

1) Eliminate by machine every record with fixed field indicating that it was taken from the MARC data base.

2) At same time, eliminate other records with language code for English and imprint date of

TABLE 4.2—*Characteristics of two high potential data bases, December 1971*

Characteristic	University of Chicago Library	New York Public Library (Research Libraries)
Size of data base	175,000	16,000 [1]
Annual growth	30,000 to 40,000	65,000
Percentage of records taken from LC MARC	17	15–20
Percentage of records in English	40 to 60	45
Percentage of nonmonographic records	2	20
Kind of cataloging data	Full entry, similar to LC	Full entry, similar to LC
LC card number present	[2]	Yes
Format	Based on ISS planning memorandum No. 3 [3]; detailed field and subfield identification.	MARC with modifications
Indication that record was taken from LC MARC	Yes	No

[1] At the time of this analysis, the NYPL system had only been in full operation for 1 month.
[2] Approximately ½ of the records do not have LC card numbers. The University of Chicago Library does not attempt to supply LC card numbers for records locally cataloged or records for which cataloging information is obtained from other than LC card sources.
[3] Avram, Henriette D., Freitag, Ruth S., and Guiles, Kay D., *A Proposed Format for a Standardized Machine-Readable Catalog Record*, Library of Congress Report, June 1965.

1968 or later. Automatic deletion of records without LC card numbers would not be desirable because the absence of a number is no guarantee that the item was not also cataloged by the Library of Congress.

3) Convert remaining records to input format for format recognition.

4) Process records by format recognition program.

5) Print records.

6) Search records against LC Official Catalog:

a) to identify other ineligible records.
b) to compare eligible records against matching LC records; for Objective 1, this involves checking only access points; for Objective 2, it involves checking the entire record.

7) Proofing for format recognition.

8) Updating to correct format recognition errors and to make changes dictated by catalog comparison.

9) Second proofing and final verification.

Essentially, the same steps would be followed for the New York Public Library records, except for the means of eliminating ineligible records by machine. Since LC card numbers are present whenever the cataloging data were taken from an LC record, there is a reasonable expectation that records without LC card numbers could be eliminated at the earliest possible stage. However, the absence of an indicator that the record was taken from MARC makes it necessary to use the language/imprint date algorithm to distinguish ineligible records among the records with card numbers.

The estimated manpower costs per record for converting these data bases are shown in Table 4.3. Also to be taken into consideration is the one-time cost of the program to convert records into the input format required for format recognition. The actual program cost per record would depend

TABLE 4.3—*Manpower costs for different conversion methods*

Function	Original conversion at LC	Conversion of other library data bases	
		Objective 1	Objective 2
Catalog comparison	$0.19	[1] $0.19	[1] $0.19
Proofing	.58	.29	[2] .58
Original typing	.24		
Typing corrections	[3] .08	.08	.16
Verifying	.59	.30	.59
Other duties [4] and leave	1.17	[5] .59	1.05
Total	2.85	1.45	2.57

[1] Base cost; the cost would increase if a significant number of ineligible records were identified at this stage.
[2] Base cost; the cost of proofing to ensure an exact match with the LC record was not estimated but would make the per record cost appreciably higher.
[3] The best available evidence indicates the typing of corrections accounts for 25 percent of the total typing cost for MARC/RECON.
[4] Includes supervision, training, and clerical activities.
[5] This cost is directly dependent on the costs of the basic functions; therefore, it fluctuates with them.

on the number of eligible records in a data base. For example, if the program costs $10,000 and the data base yields 50,000 records, the per record cost is $.20. In the case of the two data bases selected, it was assumed that the program would be useful over an indefinite period and, therefore, that in the long run the per record cost should be quite small relative to the total conversion cost.

For Objective 2, another factor must be taken into account. The number of data elements that must be added or changed affects the cost of correction. An investigation of two samples of records showed that the University of Chicago records lack a few basic LC data elements that are used in the New York Public Library records. This means that, potentially, Chicago records would be more costly to bring to the level of LC records. However, because the volume of corrections affects typing, the smallest of all manpower costs, it did not seem worthwhile to estimate the slight differences in costs that might result from conversion for Objective 2.

The costs for each of the data bases are shown as identical because the estimate of eligible records in each data base was made on the assumption that only that data base was being converted. If several data bases were converted, it can be assumed that the percentage of eligible records in each new data base would dwindle. This would have the effect of increasing the per record cost and making the per record programming cost a more significant factor.

System Considerations

Use of other data bases to increase the volume of a national bibliographic store would involve several hidden costs, not estimated in the preceding sections. These costs would relate to liaison with a number of organizations and the analysis of many file structures with varying degrees of associated documentation. Therefore, a best strategy should be developed in terms of cost as well as utility. The various data bases should be ranked according to size and bibliographical completeness with approximate estimates of proportions of eligible records (non-MARC vs. MARC records, etc.). Data base conversion should begin with the highest ranking file. However, once a few large files have been converted and put into the national store, the yield from other data bases might tend to be so low as to drive the per record cost of the conversion program too high for economic feasibility.

Any thresholds chosen at this time as to minimum size of data base and length of the record would be quite arbitrary. What is considered a threshold value would, in the end, depend on the form and/or language of the material and when the data base was being considered for conversion. For example, if a machine-readable data base of some 50,000 motion picture and filmstrip records, meeting appropriate format and bibliographical criteria, were to exist in 1975, it could be considered a candidate for a national store of machine-readable records for that form of material because the Library of Congress plans call for initiating such a service in fiscal 1973.

It has already been noted that libraries can be expected to know the characteristics of their own data bases. In terms of the national interest, it would be useful to consider establishing a standard for recording and publishing information about the form of material, the language and the content of machine-readable records in library data bases. Such a standard should simplify the chore of determining the utility of a data base and also make available to the library community as a whole detailed specifications for each individual library's data base. This standard should be developed under the auspices of the American National Standards Institute, Committee Z–39.

Data Base Acquisition Cost

The question of what charges (if any) should be made for the use of a library's machine-readable data base for a national bibliographic store must be considered. The cost of copying the file as well as that of purchasing the tapes necessary for the interchange might represent a minimum charge. Such a charge would be a fraction of a cent per eligible record. It may be questioned whether a library should recover part of its production costs in such a transaction. It could be argued that the recompense for contributing records to the national store should be measured in terms of a library's future use of the contributions of other libraries. Furthermore, contributions of original cataloging made on a continuing basis might conceivably substitute for reports to the National Union Catalog.

The situation is further complicated by the fact that several commercial firms purchase MARC tapes on a regular basis to provide cataloging services for the library community. The fact that a contributioh to the national store represents a contribution to profit-making organizations may act as a deterrent to the transfer of these files on a cost-only basis. Profit-making organizations

might acquire individual library files directly. In this event, the library could recover a substantial part of its investment by making the file available at a cost in excess of the duplication and tape purchase costs. Early release of library files to the national store could conceivably reduce the potential income to the individual libraries as some prospective buyers might await distribution of the records through the MARC service. Consideration on the part of the commercial firms will certainly be given to the length of time required for LC to process all of the data. The economic implications of this kind of data transfer should be fully investigated in the near future.

Conclusions

This study led to the following conclusions:

1) Machine-readable bibliographic data bases do exist that could be used to increase the volume of the national store. This study indicates that the per record cost of converting these records to the MARC format, comparing them with records in the LC Official Catalog, and updating their content to the point where they match those records approaches the present per record MARC/RECON cost.

2) The cost of converting the same records if only the access points were updated appears to be substantially lower than present MARC/RECON costs. The minimum cost of this method of data base conversion is probably on the order of one-half of present costs. Since these data could not be used in this form by the Library of Congress, the question of how this effort could be funded remains to be resolved.

3) Should any program be undertaken, the high potential data bases should be ranked by size and completeness of content of records. The highest ranking data base should be the first to be converted. Early consideration might be given to the University of California libraries file containing approximately 750,000 records. However, it should be noted that the character of the records would have to be evaluated to determine whether the estimated per record conversion cost held true for this data base. Lack of the necessary information made it impossible to make an analysis at the time of this study.

4) A standard should be established for reporting the form of material, language, and the content of machine-readable records in library data bases to simplify the job of determining the utility of another library's data base.

References

[1] American Library Association. Resources and Technical Services Division. Book Catalogs Committee. "Book form catalogs; a listing." *Library Resources and Technical Services*, v. 14, Summer 1970, p. 341–354.

[2] Dolby, James L. "Programming Languages in Mechanized Documentation," *Journal of Documentation*, v. 27, June 1971, p. 136–155.

[3] Aron, J. D. "Estimating Resources for Large Programming Systems." *In* NATO Science Committee. Software Engineering Techniques; report on a conference sponsored by the NATO Science Committee, Rome, Italy, 27th to 31st October 1969. [Birmingham] 1970. p. 68–79.

CHAPTER 5

On the Implications of a National Union Catalog in Machine-Readable Form

Introduction

In the simplest terms, the development of the National Union Catalog involves combining Library of Congress catalog records with those of other libraries to produce a file of discrete entries, posting to this file information about duplicate holdings in other libraries, and providing the resulting information in ways calculated to satisfy the needs of the library community. The accomplishment of this task entails many bibliographic and technical problems which are aggravated by the volume of information that must be processed to produce the end result.

The bibliographic problems relate to processing reports to build the basic file. They involve:

1) Identifying new titles.

2) Making the forms of names in main and added entries on non-LC cards compatible with names in the LC Official Catalog. (See Appendix A for a discussion of the problem of compatibility.)

3) Posting new locations to existing records. This task is classed as bibliographic because it arises from a search to determine whether a title is new to the National Union Catalog.

4) Providing necessary see and see-also references.

5) Updating bibliographic records when additions and corrections are received.

The technical problems relate to the means of disseminating information from the file. Four criteria are posited to assess the merits of the means of dissemination:

1) *Completeness:* the full bibliographic record must be given in at least one readily available source.

2) *Currency:* the information should be made available as soon as possible on a regular schedule; listings of new titles should appear at least once a month.

3) *Convenience:* both the format of the published information and the frequency of its cumulation should facilitate the work of bibliographic searching.

4) *Cost:* the cost of publishing the information should be kept as low as possible so that the National Union Catalog can be widely distributed.

It is readily apparent that considerations of cost have an important bearing on the extent to which the other criteria can be satisfied. Acceleration of the frequency of publication, broadening of the cumulation pattern, and improvement in the physical format are all directly related to the cost of producing a printed catalog. Thus, in the final analysis, decisions must be made as to whether optimizing currency and convenience justifies the cost of doing so, especially when the cost must eventually be borne by subscribers to the catalog.

The following statistics reveal the magnitude of the labor required to produce the National Union Catalog for 1970:

1) Approximately 226,000 LC catalog records were added.

2) Approximately 108,000 discrete titles cataloged by other libraries were added. Actually, a larger number of these reports were prepared for the monthly catalogs, but in the annual cumulation some were replaced by LC records issued at a later date.

3) Approximately 1,000,000 outside library reports had to be searched and a further 1,500,000, which were immediately identifiable as reports of additional locations, were forwarded for posting to the file. (See Appendix B for details on NUC reporting.)

The principal vehicle for making this tremendous mass of information available is the *National Union Catalog: A Cumulative Author List.* This book catalog is issued in what can be broadly described as monthly, quarterly, annual, and quinquennial cumulations; the actual pattern of publication will be described fully in a later section. The arrangement is by main and added name entries. Except for titles that are main entries, there is no access by title or series. Full bibliographic information appears only under the main entry; added entries take the form of references to the main entry. The main entry records consist of LC catalog cards and especially typed versions of outside library reports. Added entries and name references are typed on cards unless (in the case of new name references) a printed reference is available. All of the cards are arranged in one sequence, mounted in three columns on large pieces of cardboard, photographed in a reduced size, then printed and bound by standard methods. The 1970 annual cumulation of this catalog consisted of 14 volumes, comprising approximately 13,000 pages.

The second major component of the NUC is the *Library of Congress Catalog—Books: Subjects.* This catalog is limited to LC catalog records because the cost of editing outside library reports to provide consistent subject headings would be prohibitive. Despite this restriction, this catalog does provide partial subject access to NUC because the majority of LC items are held by other libraries. In 1970, the annual cumulation of the subject catalog required 5 volumes, comprising approximately 9,000 pages.

The third component of the NUC is the *Register of Additional Locations*, containing location reports that were received after a catalog entry has been printed in an annual cumulation. Because of the huge number of these reports, they are grouped by the year the original bibliographic record was prepared and issued in segments. Thus, the 1970 issue of the register consisted of two volumes, covering LC cards and NUC reports dated 1964 and 1965.

As the coverage of the MARC data base grows, and as the capability of local input is added at the regional level at such centers as the New England Library Network and the Ohio College Library Center, the concept of on-line union catalogs is fast becoming a reality. It does not follow, however, that knowledge so far gained from the very limited, largely conceptual, experience with machine-readable union catalogs can be extrapolated to the much larger, more complex system to provide on-line access to the National Union Catalog. It appears safe, therefore, to predict that, for some time to come, we will make use of NUC information in book form or microform. However, the growth of the MARC data base does make it feasible to produce these catalogs by computer, thereby relieving humans of much of the drudgery of preparing the catalogs and, at the same time, offering the possibility of additional access points to the bibliographic information.

Therefore, it was logical to include in the RECON studies a preliminary analysis of what would be involved in the production of the NUC from cataloging data in machine-readable form. The aim was only to consider the bibliographic and technical implications of a machine-readable NUC data base as a foundation for future investigations. The magnitude of the problems and the constraints of time, funds, and manpower available to the task force precluded formulation of a detailed system design with associated cost estimates.

Design for a National Union Catalog

Since the results of the RECON Pilot Project conducted at the Library of Congress make it unlikely that any large-scale retrospective conversion effort will be undertaken in the near future, this study concentrated on an NUC for current materials. Problems of including retrospective records and their locations were not taken into account.

In considering the optimum format for a National Union Catalog produced from machine-readable records, the RECON Working Task Force selected the register/index form of catalog because it allows favorable cumulation patterns and more points of access without having to print a full bibliographic record more than once. Basically, this type of catalog comprises:

1) A register of complete bibliographic entries arranged by numbers assigned as each item enters the system. Register volumes are issued regularly but they are never culminated.

2) Indexes providing various access points derived from the register entry. The index entry includes a brief bibliographic identification and the register

number together with any other data that may be desired. The indexes may be in dictionary form or divided into sections (e.g., name, title, subject). An index volume is issued with each register volume and the various indexes are cumulated regularly.

The future NUC, as conceptualized by the RECON Working Task Force, would have the following indexes:

1) Name index: personal and corporate names used as main, subject, and added entries (including author/title series).

2) Title index: titles used as main, subject, and added entries (including uniform title headings and series entered under title).

3) Topical subject index (including geographic subject headings).

Figure 5.1 shows the entry elements and their MARC tags for each of the indexes. Table 5.1 gives the order of data elements in each type of index entry. Index entries under added and subject entries would include the full form of main entry. Each index entry for an LC record would include the LC card number. The index entry for the main entry would include all locations reported to the date of the cumulation. Figure 5.2 presents examples of register and index entries for a typical bibliographic record.

Although index entries would be designed to be complete for many purposes (e.g., initiating an interlibrary loan, obtaining an LC card number), it would sometimes be necessary to check the register volume to obtain the full bibliographic information (as in cataloging). The disadvantage of double look-up is minimized by the fact that the second search by register number is straightforward.

A prime advantage of the register/index catalog is that each register volume is complete as issued and its contents need never be merged with those of other register volumes. The index volumes are cumulated but, as the entries are shorter, they lend themselves to compact presentation thereby effecting an overall savings in publication costs as compared with conventional book catalogs. The compactness of the indexes would also facilitate rapid scanning of entries.

FIGURE 5.1—*Entry elements (and their MARC tags) covered in proposed NUC indexes*

NAME: Entries beginning with a personal, corporate, or conference heading
- Main: 100, 110, 111
- Added: 700, 710, 711
- Subject: 600, 610, 611
- Series (traced as in note): 400, 410, 411
- Series added entry: 800, 810, 811

TITLE: Entries beginning with a title
- Uniform title heading: 130
- Romanized title: 241
- Bibliographic title: 245
- Added: 730, 740
- Subject: 630
- Series (traced as in note): 440
- Series added entry: 840

SUBJECT: Other than those included in name and title indexes
- Topic: 650
- Geographic name: 651

FIGURE 5.2—*Examples of register and index entries*

Register Entry
12345*

Ackoff, Russell Lincoln, 1919–

Fundamentals of operations research [by] Russell L. Ackoff [and] Maurice W. Sasieni. New York, Wiley [1968]

ix, 455 p. illus. 24 cm.

Includes bibliographies.

1. Operations research. I. Sasieni, Maurice W., joint author. II. Title.

T57.6.A2 001.4'24 67-27271
Library of Congress MARC

Index Entries

Name **Ackoff,** Russell Lincoln, 1919– Fundamentals of operations research. 1968.
ENG T57.6.A2 67-27271 **12345**
DLC IU CSt RP NjP

Sasieni, Maurice W., joint author. Fundamentals of operations research.
[Ackoff, Russell Lincoln, 1919–] 1968.
ENG T57.6.A2 67-27271 **12345**

Title **Fundamentals** of operations research.
[Ackoff, Russell Lincoln, 1919–] 1968.
ENG T57.6.A2 67-27271 **12345**

Subject **OPERATIONS RESEARCH**
Ackoff, Russell Lincoln, 1919– Fundamentals of operations research. 1968.
ENG T57.6.A2 67-27271 **12345**

*NOTE.—The hypothetical register number in this example is not intended to suggest the actual format of such a number.

TABLE 5.1—*Order of data elements in each type of index entry* [1]

Type of index entry	Variable elements			Invariable elements
	1st	2d	3d	
Main name [2]	Main name	Title [3]		Imprint date, Language,[4] LC call number,[5] LC card number,[5] Register number.
Added name	Added name	Title [6]	[Main entry] [7]	As above.
Title [6]	Title	[Main entry] [7]		As above.
Subject [9]	Subject	Main entry [7]	Title [6]	As above.

[1] This relates to the function of the entry, not the index in which it appears.
[2] A personal, corporate, or conference name used as a main entry.
[3] Uniform filing title, romanized title, or bibliographic title, in that order of preference.
[4] MARC language code.
[5] If present in register record.
[6] Bibliographic title.
[7] Not relevant for works entered under title.
[8] Data elements in title index entries vary considerably depending on the kind of title (uniform title heading, title main entry, title added entry, etc.). In the interest of simplicity, this line of the table describes the predominant case, entry under the bibliographic title of a work.
[9] Name, title, topic, or geographic name as subject.

Since the National Union Catalog covers the entire range of current acquisitions cataloged by the Library of Congress and the contributing libraries, most of the entries it contains are not in machine-readable form. The balance will gradually shift as funding and other resources permit the expansion of the MARC Distribution Service, and it is possible also that eventually some of the larger libraries may be able to report their new holdings directly in machine-readable form. Nevertheless, for the foreseeable future, a means will be needed to combine machine-readable records and conventionally printed records to produce the National Union Catalog.

The register volume would be made up of three types of records, each requiring different treatment:

1) A MARC record for a full LC bibliographic record converted to machine-readable form as part of the MARC Distribution Service.

2) An NUC report from a contributing library. After such a report has been certified to be for a new title and the major access points have been reconciled with the LC Official Catalog, the record would be keyed in full, processed by the format recognition programs, proofed, and verified. The keying effort is essentially the same as that required to prepare NUC copy in the present manual system.

3) LC printed cards for records outside the scope of the MARC Distribution Service.

MARC and NUC reports would be used to create part of the register by computer-controlled photocomposition techniques. LC non-MARC records would be assembled for the second part of the register using the same technique now employed for the manually produced book catalogs.

The assignment of register numbers presents certain difficulties. Machine-readable LC and NUC records could be numbered by the computer as they entered the system, but conventionally printed LC records would have to be numbered by hand. This would require a separate block of numbers for each part of the register and, to avoid confusion, the register numbers for the conventionally printed LC cards should begin with a distinctive prefix.

The indexes would be created as follows:

1) LC MARC records and NUC reports converted to machine-readable form would be processed automatically to produce truncated records for the desired indexes.

2) LC non-MARC records would be represented by a master index record in machine-readable form, which would contain all of the data elements necessary for automatic generation of the appropriate index entries for each register entry. Initially, the master index record would have to be specially produced from a printed card, but it should be possible eventually to fill this need by adapting the machine-readable record used in the projected automated LC Process Information File (PIF)[1]. In the latter case, the only additional ef-

TABLE 5.2—*Components of a national union catalog in machine-readable form*

Type of input	Type of output		Retained machine data bases	
	Register	Indexes	Register [1]	Indexes [2]
MARC	Machine-readable	Machine-readable	Register number and locations only [3].	Yes.
NUC reports	Machine-readable	Machine-readable	Yes	Yes.
LC non-MARC full record	Manual	Not available	Not available	Not available.
Master index record	Not available	Machine-readable	Yes	Yes.

[1] One file arranged by register number.
[2] One file of each type of index (name, title, subject).
[3] MARC data base is retained elsewhere for other purposes.

fort would be inputting data elements that were not required for PIF (e.g., subject entries). PIF records in nonroman alphabets would probably require special handling because the skeleton PIF record might not provide all of the data elements for the master index record.

Table 5.2 shows the type of inputs and outputs of the proposed NUC system as well as the machine-readable data bases that would be maintained.

An NUC System

A system had been hypothesized to indicate how the NUC register/index might be produced from machine-readable records. A variety of solutions could be postulated, taking into account different computers and peripheral devices. The time when such a system would be implemented, the expansion of MARC, and the state of the art of networks and regional machine-readable union catalogs would influence the design. The NUC system could be a subsystem of the LC system. If on-line access to the national node from regional nodes is in being, some or all files would be stored on random access devices. On the other hand, the NUC system could be a stand-alone system, utilizing MARC records as a source of input but maintaining its own files. If there were no requirement for on-line access by regional systems, it would be economical to design a batch processing tape system because of the large volume of data involved and the much higher costs of disk storage.

An NUC stand-alone system would receive MARC data through the MARC Distribution Service in the same manner as the LC Card Division does today. The system also would need the capability to maintain files of LC non-MARC records, NUC reports not in the LC data base, and locations. Since it is not part of this study to determine an exact method based on a detailed analysis resulting in definitive design with associated cost estimates, the following methodology should be considered as a possible way to assemble data for the proposed publication.

Since various cumulation patterns for publication of the indexes and the register of additional locations are possible, a hypothetical publication schedule has been assumed for the purpose of this description. As far as time intervals are concerned, the system is open-ended and schedules could be modified without any changes made to the system described.

The assumed publication pattern is as follows:

1) Monthly indexes at the end of each of the first two months of a quarter.

2) A quarterly index cumulation (covering the last three months) at the end of each of the first three quarters of a year.

3) An annual index cumulation at the end of each of the first four years.

4) A quinquennial index cumulation at the end of the fifth year.

5) An annual list of locations not included in the name index. The publication of location information is as follows:

a) During the first year whenever the main index entry for any given record appears (in monthly, quarterly cumulations, and annual),

all locations received to date will be printed with this entry.

b) At the end of the second year, a list of locations will appear with all additional locations for records printed in the first year.

c) All locations for this record received after the second year will be cumulated and appear in the quinquennial index main entry.

d) Additional location reports for older records not in a quinquennial index will be published in a separate list with the quinquennial.

The LC MARC records and NUC reports would be used to generate that segment of the register produced automatically. At the same time, a unit card is produced to be filed for searching reports from outside libraries against all NUC entries to identify new entries and to post locations. The LC non-MARC records are used to produce the manual segment of the register and a copy of the record with the assigned register number is sent to an input section for keying the master index record. For the remainder of this section, the machine-readable record derived from the LC non-MARC record will be referred to as the non-MARC record. It should be kept in mind that this record is not a complete bibliographic record. LC printed cards representing these non-MARC records are also filed for searching purposes. Since a large proportion of the NUC reports are for the retrospective entries, this search file would be maintained even if all current entries were searchable in a machine mode.

The MARC records are the machine-readable Library of Congress bibliographic files. Since LC is responsible for the MARC Distribution Service, these records are organized in such a manner that the date of last transaction [2] is readily available for the purpose of distributing new, corrected, and deleted records (by status of record code) during a prescribed period of time. This same type of control on date and status is required for the publication of the register/index. Therefore, the NUC reports and the non-MARC files would have to be organized to allow this capability.

The register is published monthly and register entries are never reprinted. Between publications, bibliographic records are corrected and/or deleted when necessary and new bibliographic records are added to the file. In addition, an NUC report can be replaced by an LC record (MARC and non-MARC). Since a library can report a holding to a published record at any point in time, and the library may not be aware that an LC record has replaced the reporting library record, a reference is made from the number of the NUC report to the LC record in the machine-readable file or in the manual file, if one is maintained.

What is being produced is an updated version of each machine file composed of the following:

1) All records which have required no updating.

2) The updated form of records to which additions or changes have been made.

3) Records on the file which have been flagged as deleted.

4) All new records input since the publication of the last register/index.

Any updated record becomes a new entry with a new register number and this record is published in the next print cycle of the register (including LC records which have replaced NUC reports). When the next cumulative index listing is published, the new register number is associated with the index entries for the original record. Therefore, there is no longer any index entry pointing to the supplanted register entry. When a bibliographic record is deleted from the machine-readable data base without being replaced by another record, the index entries are deleted from the next cumulative index listing.

Both the MARC and non-MARC records have LC card numbers and NUC reports are assigned an NUC number with similar characteristics. Therefore, the LC or NUC card number [3] is used as a two-way link between the file of NUC locations and the bibliographic files. When a bibliographic record is entered in the machine file, a location record is generated using the card number and the NUC symbols for libraries holding that title. The location file is organized in such a way that it is possible to date any action taken in relation to it. The deletion of a bibliographic record would automatically cause the deletion of the associated location record or its transfer to the card number of a substituted record.

Depending upon the requirements for these records beyond printing the NUC, the location file may reside on disk or be maintained on tape. The index records must be maintained for the production of cumulative index publications and in this form the main index entry contains location infor-

TABLE 5.3—*Input and output files in the NUC system*

INPUT		OUTPUT
Bibliographic file	*Location file*	*Interim file*
1. New record added during time period	Original and added locations	Added [1] bibliographic record with locations appended.
2. Updated record updated during time period	Original and added locations	Updated bibliographic record with locations appended.
3. New or updated record from previous time period not to exceed one year.	Locations deleted or added during time period.	Bibliographic record with updated locations appended.
4. Deleted bibliographic record during time period.	----------	Deleted bibliographic record.
5. New or updated record from some previous time period exceeding one year.	Locations deleted or added during time period.	Bibliographic record with updated locations appended.
6. ----------	Locations deleted or added during time period.	Updated locations.

[1] The underscore indicates the status given each record for updating the indexes.

mation. A system could be designed to maintain only those location records which will appear in any future print cycle of the register of additional locations, provided the only requirement of the system is the publication of the register/index and the register of additional locations.

Each month, new and updated MARC records and NUC reports for the period are automatically assigned a register number and output [4] for publication of the register. Non-MARC register entries with preassigned register numbers are published in a manual mode at the same time. During this processing cycle, the following actions are performed and the resulting records written to an interim file:

1) Each new or updated bibliographic record is passed against the location file and the location record appended.

2) For any bibliographic record containing a delete status code, the associated location record will have already been deleted from the location file and therefore this bibliographic record enters the system without an appended location record.

3) Likewise, each location record residing on the location file within the time span for the publication of the indexes (i.e., added locations to a bibliographic record previously printed in an index) causes the selection of the associated bibliographic record for reprinting in the next issue of the index. Those location records that have no corresponding bibliographic records in the machine file (because they refer to records in the manual NUC) are selected for eventual publication in the list of additional locations. A location record for a bibliographical record for a prior year is also selected for inclusion in the list of additional locations. In these cases, however, the locations are appended to the associated bibliographic record for later inclusion in the quinquennial index.

Table 5.3 shows the status of the records concerned with the bibliographic files and the location file at the time the register is produced and as the record enters the indexing subsystem.

The bibliographic records are used to generate the name and title indexes and, in the case of LC records, the subject index. Locations are appended to the applicable index entry; that is, to main entries in the name index and to title main entries in the title index. Records with locations only are carried along for later inclusion in the list of additional locations. Each index entry will be written onto its own output tape. Each tape will then be sorted on the key [5] appropriate to it, i.e., names, titles, and subjects.

Since the cumulative index files are maintained in sort-key order, each updated record for a data element that is used as the major filing element must have both the incorrect version and the updated version. The incorrect version is used to find the record on the file that will be replaced by the updated version of the record. Therefore, to correct a record, the system generates from each updated bibliographic record, a delete and add record combination (two records); to delete a record, the system generates a delete record only; and to add

a new record, the system generates an add record only.

When a data element that is not used as a filing element is to be corrected, only a replace record need be generated. In this content, a replace record is a single record that causes a previous record to be deleted. However, it may prove simpler for consistency of software to treat all corrections the same. Therefore, in the case where only the locations have been affected, the system also generates a delete-and-add combination.

Maintaining the index files in sorted order should significantly reduce computer processing time since the new index file requiring sorting is relatively small, and the merge operation, a far simpler and less time-consuming procedure, is executed by incorporating the smaller file into the larger cumulative file.

There are four updates in the machine system: monthly, quarterly, yearly, and the final (quinquennial). Each update produces index files and, where applicable, the list of additional locations, and accumulates the records to be passed along for the next higher accumulation period (except for the final which only produces the quinquennial indexes). The update pattern is shown in the following schematic diagram:

M=monthly index file
Q=quarterly index file
Y=yearly index file
Numbers=months, quarter, year concerned; zero is used for location reports for records prior to M1; that is, those in the manual NUC catalog.
n=new bibliographic record
u=updated bibliographic record
d=deleted bibliographic record
L=location record (L is used both for posting new locations or deleting locations from bibliographic records in the machine as well as the manual system. In reality, location records referring to bibliographic records in the manual system would have to be indicated as such.)
[ ]=records carried in the system for the appropriate cumulation period and not printed in the month, quarter or year where they are enclosed in brackets.

The numbers always indicate the month in which the original new record was entered into the system. Thus, "u 1–23" means all records from those months that were updated in the current month (month 24). Once an update or delete action has been taken, these transaction records no longer are retained in the system

FIGURE 5.3—*Schematic representation of machine files required for the NUC system*

M1 =n1+[LØ]
M2 =n2+u1+[dl]+[LØ]+[L1]
M3 (Q1)=n1–3+u1–2+dl–2+[LØ]+L1–2
M4 =n4+u1–3+[d1–3]+[LØ]+[L1–3]
M5 =n5+u1–4+[d1–3]+[d4]+[LØ]+[L1–3]+[L4]
M6 (Q2)=n4–6+u1–5+[d1–3]+d4–5+[LØ]+[L1–3]+ L4–5
M7 =n7+u1–6+[d1–6]+[LØ]+[L1–6]
M8 =n8+u1–7+[d1–6]+[d7]+[LØ]+[L1–6]+[L7]
M9 (Q3)=n7–9+u1–8+[d1–6]+d7–8+[LØ]+[L1–6]+ L7–8
M10=n10+u1–9+[d1–9]+[LØ]+[L1–9]
M11=n11+u1–10+[d1–9]+[d10]+[LØ]+[L1–9]+[L10]
M12(Y1)=n1–12+u1–11+d1–11+LØ [1]+L1–11

.
.
.
.

M24(Y2)=n13–24+u1–23+[d1–12]+d13–23+[LØ]+ L1–12 [2]+L13–23

.
.

M36(Y3)=n25–36+u1–35+[d1–24]+d25–35+[LØ]+ [L1–12]+L13–24 [3]+L25–35

.
.

M59=n59+u1–58+[d1–57]+[d58]+[LØ]+[L1–57]+ [L58]
M60(Y5)=n1–60+u1–60+d1–60+LØ+L1–60

[1] List of additional locations for manual NUC catalog.
[2] List of additional locations for year 1.
[3] List of additional locations for year 2.

All files are sorted prior to publication and/or merged into the next higher level accumulation in the following descending sort hierarchy:

Sort key
LC or NUC card number
Date
Delete flag
Add flag

This ordering brings together in date order entry deletions and additions for a given unique index entry (sort key, card number) for the merging process.

Prior to the actual publication of the indexes, it would be necessary to pass the files against computer-based authority files to add the reference information to the name and subject indexes. Since a record which represents a new title but repeats a name or subject used in the last published edition can be added to the files, it would also be necessary to remove duplicated references from the files

479–312 O—73——3

and to insure against the inclusion of any blind references in the index prior to publication.

Using the NUC figures for the years 1966–1970, the number of index records and cross references that would be generated and their record lengths were estimated from available statistics at LC as well as LC MARC statistics produced by Columbia University. Assuming magnetic tape with a density of 1600 cpi and a blocking factor of 20, the quinquennial indexes would require a total of approximately 34 tapes (17 for name index, 9 for title index, and 8 for subject index).

Cost Factors

The maintenance of the National Union Catalog and the publication of its holdings entail many functions. In the present system all require manpower; in the projected system many of these functions would be performed all or in part by computer. Although estimating actual costs for an automated NUC is beyond the scope of this study, it seems worth considering how these cost factors would be affected by a change from one mode of operation to the other.

Editorial Cost Factors.

The effect of automation of the NUC on each of the major editorial functions is discussed in the following paragraphs. For details about the specific duties in each of these functions, see Appendix C.

1) *Arranging and Sorting.* Hundreds of thousands of LC cards and NUC reports are assimilated into the NUC data base each year. At the outset, they are received, recorded, and sorted by hand for further processing. In view of the volume of activity and the high proportion of overlapping reports, there is no reasonable expectation that it would be economical to automate this function. Therefore, the cost of this function would remain essentially the same in the proposed automated system.

2) *Searching.* The searching function constitutes one of the major cost factors in maintaining the NUC, but it does not appear to be one susceptible to amelioration by automation. The preponderance of NUC reports are duplicates of entries already in the system and thus they become merely reports of added locations. Although many NUC reports of added locations are submitted in the form of LC cards, at least as many are not so readily identifiable. Given the variation in critical data elements in many reports, it is difficult to conceive of an effective machine searching technique that would not involve excessive keying to secure an exact match frequently enough to make the process worthwhile. Moreover, even if machine searching were practical, it would be many years before the machine data base was large enough to satisfy a reasonable proportion of the searches. Therefore, it cannot be expected that the automation of the NUC would have any effect on the cost of this function.

3) *Editing.* Most of the tasks comprised by the editorial function will be unaffected by the proposed automated system. The editing of NUC reports for new titles constitutes the major workload and there is no likelihood that this task can be lightened by the computer until name authority information is available in machine-readable form. However, some cost reduction will be possible because certain editorial work in providing added entries and references will be superseded by the automatic generation of these entries from machine-readable bibliographic and reference control records.

4) *Keying.* Typing NUC reports for conversion to machine-readable form would involve approximately the same effort as typing these records in the manual system. Typing the master index record for the LC non-MARC records would entail a workload that is not required for the main entry in the manual system. On the other hand, typing added entries and references would be unnecessary for all types of records since these access points would be generated by the computer. Thus, the overall cost of typing should be somewhat lower in the automated system.

5) *Proofing.* Proofing of added entries and references would no longer be necessary in an automated system because these data would have already been proofed as part of the creation of the original record from which they were generated. In the case of LC non-MARC records, it is assumed that the added difficulty of proofing the master index record would be offset by the fact that proofing separate added entry records would be unnecessary. The proofing of the NUC register record would become somewhat more difficult because it would involve also verifying the accuracy of the format recognition processing. Across the board, however, some reduction in the cost of proofing may be anticipated.

6) *Filing.* Once the records are in the MARC format, it is no longer necessary to arrange entries by hand for the book catalog indexes and even filing the manual control file would be facilitated by the machine sorting of new entries prior to actual filing. Thus, a substantial reduction in the cost of this function could be expected.

7) *Mounting and Stripping.* The present manual method of mounting printed cards for reproduction and then stripping them for later cumulation would be unnecessary for any entry in machine-readable form. Such an entry would be processed by a computer-driven photocomposition device and cumulations would be produced by machine without regard to the printed form of earlier issuances. Even in the case of LC non-MARC records there would be some saving because there would be no need to strip the cards for cumulation.

8) *Supervision.* Since the cost of supervision tends to be a relatively stable percentage of the aggregate cost of other functions, it may be assumed that a reduction in those costs would produce a corresponding decrease in the cost of supervision. It is not possible to estimate, however, whether this reduction would be significant because the complexities of the automated system might make greater demands for supervisory time.

Noneditorial Costs.

The primary noneditorial costs are those involved in printing and binding the issues of the catalog. They are influenced by such factors as: 1) the number of times an entry must be reprinted in the course of various cumulations; 2) the amount of information included in each entry and the resulting number of entries that can be fitted on a page; and 3) the form of the hard copy.

Under present practice, the full entry may be printed as many as four times: in a monthly issue, a quarterly, an annual, and a quinquennial. However, because entries with imprints falling outside of the current three-year period do not appear in monthly issues and because the fourth quarterly is not published at the end of the year, the average entry is printed 3.24 times. The register/index catalog has the advantage of requiring that the full entry be printed only once.

Since the indexes are cumulated, index entries do appear more than once in the life cycle of the publication. However, the index information contains only those elements required to facilitate the use of the index entry as a stand-alone entry in addition to providing a link to the full entry in the register. Due to this reduction in the content of the entry, many more entries can be printed on a page. The number of entries in the present three-column format of the NUC is 27 per page; the estimated number of entries per page in a three-column format for the name index is 81.

The cost of publishing the register, the various indexes and the register of additional locations is also dependent on the form of output selected. As all information (with the exception of the full non-MARC entries) would be in machine-readable form, several main alternatives are available:

1) Graphic arts quality through a photocomposition device.

2) Reduced quality through Computer Output Microfilm (COM) to lithoplate.

3) Microform.

For each option the cost of publication is further dependent on the cumulation schedule chosen. In the present manual system, *Books: Authors* is published monthly, quarterly, and annually; *Books: Subjects* is currently published only quarterly and annually. In the proposed system it has been assumed that the name, title, and subject indexes would be published monthly, quarterly, and annually.

In the present manual system, noneditorial costs (i.e., printing, binding, shipping, etc.) account for nearly half of the total NUC cost. While it is beyond the scope of this study to identify and evaluate the many combinations of forms that are possible, it is evident that significant savings could be effected by using microform for the indexes. A conservative approach to this means of cost reduction would be to issue monthly indexes in microfiche and the quarterly, annual, and quinquennial cumulations in conventional print form.

The anticipated effect of automation of the National Union Catalog on the costs of various functions is summarized in Table 5.4. The significance of the variations is difficult to assess because the various functions do not contribute equally to the total cost and the summary of editorial costs does not take account of the cost of record control procedures that typify complex machine input operations. Therefore, although the overall cost of the proposed automated system may be less than that

TABLE 5.4—*Increase or decrease in the cost of producing the* National Union Catalog *by computer relative to the present manual method, by function and type of record* [1]

Function	LC MARC		LC-non-MARC		NUC report	
	Register	Indexes [2]	Register	Indexes	Register	Indexes
Editorial:						
Arranging, sorting	−	−	=	−	=	−
Searching	NA [3]	NA	NA	NA	=	=
Editing	NA	−	NA	−	=	=
Keying	NA	−	NA	+	=	−
Proofing	NA	−	NA	=	+	−
Filing	−	−	−	−	−	−
Mounting, stripping	−	−	=	−	−	−
Supervision	=	=	=	=	=	=
Noneditorial:						
Printing	−	=	−	=	−	=
Binding	−	+	−	+	−	+
Shipping	−	=	−	=	−	=

[1] Relative cost shown by following symbols: + (greater than manual cost), = (same as manual cost), − (less than manual cost).
[2] Added entries in the manual system are analogous to index entries in the machine system.
[3] Not applicable; used when a function is not necessary in either system.

of the manual system, it would be more prudent to say that it should not exceed that cost.

A Model NUC Network

An NUC reporting system could be organized on the basis of regional bibliographical centers that played an intermediary role by coordinating reports from their areas and by helping area users to obtain desired material. At a highly developed stage, such centers could be responsible for 1) scrutiny, verification, and possible alteration of incoming records as an initial step in their integration into the NUC file, and 2) referral of requesters to material or inter-library borrowing and lending operations in response to either search requests from within each region or queries transmitted from among those received at the national center. Each of these functions will be considered briefly.

Most of the tasks connected with integration of data into the NUC store could be performed at the regional level, subject to the completeness and currency of authority records maintained at the regional centers and the capability of the centers to manipulate data in machine-readable form. However, the final reconciliation of incoming records with the LC Official Catalog would continue to be a task of the national agency unless it were possible to have the entire LC Official Catalog in on-line mode at locations throughout the country. The following functions might be assumed locally:

1) Development of subject and/or form responsibilities for specific libraries within each region to channel reports of locations for particular items.

2) Coordination and periodic transmission to the national center of location reports for items already known to be cataloged by LC and in the MARC store.

3) Coordination and periodic transmission to the national center of location reports for items already known to be cataloged by LC but not designated as being in the MARC data store. If records for these items have already been encoded in machine-readable form at the local level (or if the capability to convert manual records exists at the regional center), they might be converted to the LC MARC format by the processes described in Chapter 4.

4) Coordination and periodic transmission to the national center of data and locations for items not already known to be cataloged by LC. In such cases it is likely that a division of functions between the national and regional centers would be desirable. At the least the regional centers would coordinate the reporting of locations on the basis of assigned responsibilities for coverage and indicate to the national center whether or not cataloging practices conformed to those of the Library of Congress. Further action toward integration into the NUC file might be possible as facilities and available data at the regional centers expanded.

In a network of regional centers it is not clear whether locations would best be reported in the NUC outputs as those of specific institutions or simply as items held within particular regions. The former procedure would allow for the continuation of direct referral of a search request to a library holding the item, although the requester would not necessarily be informed of other locations within the region which might be more advantageous for referral purposes in given instances. Although reporting in terms of regional centers would require the center to act as a "middleman" in the handling of search requests, it

might allow for greater flexibility and rationality in the flow of requests for items to be borrowed. The choice of reporting schemes would depend on what capabilities the regional centers developed and what roles they were willing to assume.

Conclusions

Automation of the National Union Catalog using the register/index form would have the following advantages:

1) The range of access points to the bibliographic data would be extended to titles and series.

2) All types of indexes would be cumulated and published on the same schedule.

3) The time required to produce cumulations would be significantly reduced.

4) The cost of the automated system offering these advantages for monthly, quarterly, and annual issues would not exceed the cost of the present manual system. The cost of producing the quinquennial would be sharply reduced.

5) The cost of the automated system should gradually be reduced as more languages are covered by the MARC Distribution Service. Further cost reductions may be possible as other libraries are able to report their holdings in machine-readable form.

6) Converting NUC reports and master index records for LC non-MARC records to machine-readable form would create a data base that could be searched by nonconventional access points (e.g., language, imprint date, geographic area).

7) The NUC data base might eventually form the nucleus of an on-line network of regional bibliographic centers.

References and Notes

[1] This project is described in Avram, Henriette D., Lenore S. Maruyama, and John C. Rather. "Automation activities in the Processing Department of the Library of Congress." *Library Resources and Technical Services*, v. 16, Spring 1972, p. 195–289.

[2] The date of last transaction for new records would be the same as the date entered on file. For modified or deleted records, the date of last transaction is the date that any processing was performed affecting a particular record and therefore the record must again be distributed to subscribers.

[3] The register number could be maintained as the link to the location index but using the LC card number allows direct access to locations when that number is known from a source other than one of the indexes.

[4] Output in this context means formatting the records and writing them on magnetic tapes as input for a photocomposition or COM device.

[5] The sort keys will be computer generated by a program designed to satisfy the filing requirements of the published indexes.

CHAPTER 6

Alternative Strategies for RECON

Introduction

Experience in the RECON Pilot Project indicates that it would be impractical to undertake the large-scale conversion project envisaged in the original RECON study. On this scale, such a project would demand far more staff, space, and money than there is any reasonable prospect of obtaining. A retrospective conversion project on a lesser scale has the evident disadvantage of being too slow in responding to the needs of individual libraries aiming toward automation involving total conversion. It appears to be a fact of life that many libraries are disinclined to postpone local efforts until records are available from a central source. Therefore, the library community is still faced with costly conversion efforts resulting in multiple files of nonstandardized data as well as duplication in titles converted.

For these reasons, the RECON Working Task Force felt the need to reexamine the premises of its original study to determine whether an alternative strategy might offer a better prospect of satisfying the need for retrospective conversion. The present chapter considers the merits of systematic versus nonsystematic conversion as well as alternative ways in which the records might be made available.

In attempting to evaluate the advantages and disadvantages of various strategies, the Working Task Force was constantly faced with the realization that there is no perfect solution to the problem. The critical questions of the languages to be covered, the dates of the records, the forms of material, the extent of the bibliographic information, and the details of the machine format yield widely different answers depending on the type and size of library involved. Therefore, the best that can be hoped for is a compromise on the requirements of libraries of various types and sizes. The ensuing discussion is an attempt to reach an optimum solution to the problem.

Systematic versus Nonsystematic Approach

In the context of this discussion, systematic conversion means the orderly conversion of existing LC records by date and language. This allows a potential user to predict with reasonable certainty whether a desired record is in the data base.

The systematic approach to retrospective conversion recommended by the RECON Working Task Force has the advantage of offering a full MARC record of the quality of the LC Official Catalog and a clear definition by date and language of records that are in machine-readable form. It is obvious, of course, that from the standpoint of any given user systematic conversion has the disadvantage of requiring a long waiting period before all relevant records are available.

Nonsystematic conversion applies to conversion of subsets of existing records that are defined by less precise criteria; for example, all records represented in a bibliography. In such a case, a potential user can determine whether the record is available in machine-readable form only by checking the bibliography in question or by querying the data base. The conversion of records from another library's data base has this same disadvantage; namely, that there is no easy way to tell whether a specific record has been converted.

Systematic conversion of retrospective records by year of card series and language can be shown to be inadequate even to meet the needs for current acquisitions. An analysis of LC card orders for a one-year period shows a remarkable demand for older records. While it is true that 79 percent of the total number of card orders were for titles published in the last 11 years, the fact remains that 52 percent of the titles ordered were older

than 11 years (see Appendix D). The analysis of titles ordered once shows a striking consistency in the demand for uncommon titles; the percentage of single orders for titles in the latest series is scarcely different from the corresponding percentage in the oldest series. It may be assumed also that a substantial proportion (perhaps even the majority) of the titles ordered once this year will *not* be ordered at all next year and that they will be replaced by titles that were inactive this year. Thus it seems that, because of the pattern of current acquisition of retrospective materials in American libraries, a substantial body of retrospective records would have to be converted even to meet current demands for machine-readable records.

An alternative approach to RECON would be to undertake the conversion of titles ordered more than a specified number of times (say, more than 3) on the assumption that a retrospective title being acquired currently by that many libraries is likely to be held by many other libraries. Even with this approach, however, the number of records to be converted would be very large (in the specific case, approximately 425,000 records) and the coverage of titles needed by any particular library would necessarily be incomplete. On the other hand, this approach has the advantage of resolving the problem of selecting records that would satisfy the largest number of libraries of various types and sizes.

Alternative Forms of Conversion

Regardless of the data base chosen for conversion, it is necessary to settle the question of the form it will take. The RECON feasibility study recommended conversion of the full bibliographic record to machine-readable form. It would be possible alternatively to create machine-readable indexes to the data base and to store the full records in microform. A variation of this possibility would involve producing the index records and relying on the printed National Union Catalog as the source of the full records.

The cost of putting the full record in machine-readable form varies with the source of the data and the extent to which they are made consistent with the LC Official Catalog. The range is from $2.85 for an LC record to $1.45 for an outside library record for which only the major access points have been verified (see Chapter 4).

The concept of the index entry in lieu of the full record entails a basic dilemma. The more data elements included to make the index entry self-sufficient, the more the cost of creating it tends to approach the cost of a full record. On the other hand, as data elements are eliminated in the interests of economy, the index entry becomes progressively less responsive to various bibliographic needs. In the latter case, truncation of the record has the effect of severely limiting the library functions that can be completely automated by using the record. For some purposes, the need could be met by consulting the full record in another source (e.g., microform or book form) but the trade-off between economy of machine input and cost of human effort in use may be difficult to evaluate.

It was such considerations as these that convinced the Working Task Force to recommend in its original study the conversion of the full bibliographic record to the MARC format, and to confirm that conclusion in its study of levels of machine-readable records (see Chapter 3). The advantage of having a full MARC record for national purposes is that, regardless of the intended use, the required information is available.

A factor to be considered in evaluating the merits of a system involving a machine index to a microstore of full bibliographic records is the cost of maintaining the microstore. Existing equipment for storing large numbers of microimages seems always to be expensive, especially when it must be capable of providing relatively rapid access to individual microimages. Another disadvantage in any proposal to use this technique on the national level is the procedural complexity of implementing it. The problems of which file should be filmed, how it would be filmed, and how the index records could be efficiently created from the source data should not be underestimated. They are in fact the same problems that were discussed in connection with the microfilming of records in the RECON Pilot Project.[1]

In the case of creating an abbreviated machine record and relying on the existing NUC book catalog for the record, the present difficulty in locating a particular entry, especially revised entries, among the various alphabetic sequences of NUC would remain. This disadvantage could be lessened by including in the abbreviated record (at additional cost) a number for the NUC volume containing the full record. Experience in the RECON Pilot Project suggests, however, that the difference in cost between an index record and a full record would not be sufficient to offset the difficulties (that is, the costs to the user) of obtaining the full record when it was needed.

Conclusions

In the light of the foregoing considerations, the RECON Working Task Force feels the large-scale retrospective conversion should be undertaken by a centralized agency (or component of an agency) established expressly for that purpose. This effort should not divert the Library of Congress from its present objective of going forward as rapidly as possible to convert all of its current catalog records to machine-readable form. To the extent that retrospective records are required for Library of Congress purposes (e.g., Card Division mechanization; special book catalogs), LC would convert these records according to its present practices. The central agency should have two major functions:

1. It should undertake a program to convert the retrospective LC records that are most in demand. Initially, the criterion for selection might be those records ordered from the LC Card Division more than a specified number of times.

2. It should be responsible for adapting machine-readable records from libraries other than LC. The scope of this cooperative approach would be modified as each new language is covered at LC.

In developing its program and carrying out these tasks, the agency should draw on the experience gained in the MARC and RECON activities at the Library of Congress. Since users will be obtaining current catalog records from the Library of Congress, it is essential that the products of these two enterprises be entirely compatible.

To ensure that the conversion of other libraries' machine-readable data bases result in consistent records, the following procedures are recommended:

1. If a library converts, it should use the best available LC record.

2. If at all possible, the full MARC format should be used.

3. The centralized agency should undertake to process records to bring them to the full MARC format (if necessary) and to make the access points compatible with the LC Official Catalog (see Chapter 4).

The question of how such an agency could be funded is beyond the scope of this study. Since the heavy expenditure involved would have to be justified in national terms, it seems reasonable to suppose that the operating expenses of the agency might come from Federal sources. It is possible, however, that foundation funds could be obtained to underwrite the costs of planning the organization and supporting it during a test period. The investigation of these possibilities might be an appropriate task for the National Commission for Libraries and Information Science.

Reference

[1] RECON Pilot Project. RECON Pilot Project; final report. Prepared by Henriette D. Avram. Washington, Library of Congress, 1972, p. 39–43.

APPENDIX A

Problems in Achieving a Cooperatively Produced Machine-Readable Bibliographic Data Base

*by Paul B. Kebabian**

In assessing the utility of a machine-readable bibliographic record, the feasibility study prepared by the RECON Working Task Force in 1969 stated:

> A prime reason for converting catalog records to machine-readable form is to achieve greater flexibility in manipulating data. This flexibility will facilitate searching and retrieval; it will lessen the effort of updating records; and it will contribute to production of a wide variety of cataloging products (cards, book catalogs, special lists, book labels, etc.). Although initially most of the applications will be along traditional lines, computerization of cataloging data should give an added dimension to bibliographic control that may materially alter familiar patterns of use.[1]

In the following remarks, these *a priori* assumptions are made: 1) the development of a machine-readable bibliographic data base, consisting of retrospective library catalog records which can be acquired, or to which access can be made by many libraries or groups of libraries, is a desirable objective; and 2) the reasons why such an achievement would be of great value to library service, as stated in the original RECON report, are essentially valid.

In essence, the problem of achieving a bibliographic data base by cooperative means is related to the nature of the record. By "nature" I refer to the characteristics of the record in terms of its constituent elements as defined and prescribed by cataloging codes and standards of practice for the order and content of the catalog entry, the subject terminology, and classification. The systematic application of codes of principles and practice, authority lists, and standardized classification schedules in preparing a bibliographic record is desirable if not essential for maximum utility and accessibility. This need obtains whether the end product is a cooperatively produced machine-readable catalog record, a traditional card form union catalog, or the catalog of an individual institution.

In considering the scope of a project to convert retrospective bibliographic records to machine-readable form, the RECON Working Task Force report proposed that first priority be assigned to English language monographs from 1960 to 1969, followed by Romance and German language monographs from 1960 to 1969 and English language monographs from 1898 to 1959.[2] The question of the records to be converted had another major dimension, namely the source or sources from which the records would be drawn.

Several existing card form, book form, and machine data bases are obvious possibilities. They include the National Union Catalog, existing regional union catalogs, the catalogs of a selected group of major research libraries, the Library of Congress Official Catalog, and the computerized catalog records of institutions or combinations of libraries that have already converted files as part of their automation applications. If large-scale retrospective conversion is a desirable end, then the maximum desideratum would seem to be the largest master file available. The Library of Congress cataloged some 4.2 million titles in the period 1898–1969.[3] The National Union Catalog (NUC) consists of an estimated 11 million titles, including

---

*Mr. Kebabian is director of libraries at the University of Vermont. He was formerly associate director of libraries at the University of Florida and chief cataloger at the New York Public Library.

the LC entries. The breadth of coverage of the NUC is large in comparison with the holdings of the major regional union catalogs. In 1942, it included over 80 percent of their holdings, but none of the regional union catalogs had more than 9.2 percent of the NUC titles.[4]

The RECON report took into consideration a variety of sources to serve as a possible base for catalog record conversion and concluded that the most satisfactory base would not be the NUC, but rather the LC Official Catalog. For technical reasons, however, conversion would begin with cards from the LC Card Division "record set," a file containing a master copy of the latest revised reprint of each LC catalog card. The reasons for this conclusion are discussed in the RECON report.[5]

Perhaps the most realistic and compelling reason for this choice was the recognition that, if the conversion project was to result in a useful product offering the potential for a variety of applications, the data base should be derived from the source offering the greatest consistency and standardization in its bibliographic information. Although there may be no positive evidence in the form of studies of consistency of cataloging standards observed by the Library of Congress over the years, empirical evidence does exist in the LC card and book catalog products.

At the same time there is evidence that other libraries have observed varying local cataloging standards. This information is provided by studies of changes made in main and added entries, in subject headings, notes, and classification on LC cards used in other libraries. A study by John Dawson [6] analyzed the kinds of changes made in 2,679 LC cards by nine major university libraries using LC cataloging copy. It revealed that less than half of the LC cards used were incorporated in catalogs without change. Although main and added entry changes were proportionately low, libraries using the LC classification changed 15.55 percent of the numbers. On 15.45 percent of the cards, the LC subject headings were either altered, supplemented, or not used.

Other evidence of a lack of consistency is provided by a cursory examination of outside library entries in any part of the printed NUC. Johannes Dewton, writing in 1961 about the draft of the cataloging code then in process of development, observed: "1. That under the present Cataloging Code there is a considerable lack of uniformity of cataloging . . . especially in the field of corporate authorship [and] 2. That uniformity is desirable, even needed, in order to exploit to the best advantage the resources of American libraries . . . and the possibility of machine control of information makes this uniformity a focal point of interest.";[7] Mr. Dewton was reflecting on a lack of standardization chiefly in the area of main and added entries as they affected the card form and published National Union Catalogs. He provided 60 examples to illustrate inconsistencies as submitted to NUC by "significant research libraries."

An important effort in the cooperative preparation of bibliographical data was the LC cooperative cataloging program which, at its peak involved participation of over 150 American libraries. It was initiated in late 1932 under sponsorship of the American Library Association Cooperative Cataloging Committee and the Library of Congress with the subsequent assistance of a grant from the General Education Board. In the initial twelve-year period, 1932–1943, the Library edited and printed catalog cards for some 96,000 titles submitted by cooperating libraries.[8] Fourteen years later, Dawson stated that "cooperative cards make up over one third of the LC cards used by research libraries for foreign-language titles." [9]

Difficulties of handling cooperative copy for printing of cards were recognized and commented on early in the program. In 1934 Charles H. Hastings, Chief of the Card Division, noted: "The item of cooperation with outside organizations that has given us most concern and has drawn most heavily on the time and energy of the division has been the revision and the proofreading of the entries supplied by libraries that are cooperating under the direction of the A. L. A. Cooperative Cataloging Committee in the cataloging of series and books in foreign languages. As anticipated in my report for last year, these entries have proved difficult to handle because nearly all are in foreign languages, and they bring up many unsettled points in cataloging, difficult to handle by correspondence." [10]

Again in 1941 following the establishment of the Cooperative Cataloging Section in the Descriptive Cataloging Division at the Library, it was noted that ". . . an attempt was made to bring the cooperative cataloging more in harmony with the Library of Congress work and to make more use of the cards produced in the cataloging of the Library's own books. Previously, some of the catalogers at the Library of Congress held such a low opinion of the cooperative cards that they often ignored them when the book was received in the Library, and did the work again." [11]

By 1967, cooperative titles edited for other libraries had dropped to 2,295 titles.[12] In 1968 with the "shared cataloging" project initiated under provisions of Title IIC of the Higher Education Act of 1965 well under way, contribution of cooperative copy ceased. With shared cataloging as a centralized activity at the Library, the opportunity for maximum standardization of cataloging records does exist because the cataloging product emanates from a single source. "One of the most significant future implications of the present [shared cataloging] program is the possibility of achieving greater bibliographic compatibility," James Skipper remarked in reviewing the project.[13]

The development of a data store of retrospective cataloging records from a number of contributing sources comes up squarely against problems of standards, uniformity, and compatibility, whether the sources be traditional card or book form catalog entries or machine stored data. The reasons for the dilemma are not difficult to perceive.

First, cataloging at any one institution is performed in relation to the body of cataloging data which it has developed through the years of its existence and incorporated into its own cataloging record. Second, the cataloging product is governed by codes specifying guiding principles and rules of practice, authority lists of subject terms, classification schedules, standardized lists of names (personal, corporate, and geographic), and similar criteria of authority. The final record is also influenced by human judgment and competence. All of the cataloging criteria have been in an evolutionary process over the years and are subject to future changes. How consistently libraries have applied codes and other criteria and how extensively they have modified prior data to reflect changes are open questions. The published NUC suggests that much inconsistency and few changes (other than revision and editing of main and some added entries) have been introduced in outside libraries.

A small sampling of NUC entries provided by contributing libraries quickly brings into focus the critical problems of compatibility among name entries, subject headings, and classification. All of these elements would be vital for successful machine processing of a full bibliographic record for the following purposes: 1) to search and produce catalog card sets, 2) to search by topical subject terms and personal or corporate names used as subjects, 3) to identify records by classification number, and 4) to search by author and title. Other data elements encoded in preparing the record might also be used for search and print-out, visual display or other retrieval capability as well as for uses only vaguely perceived at this time.

The Sterling Memorial Library at Yale includes a major collection of literature in German and Romance languages. In addition, Yale has cataloged thousands of dissertations of continental scholars, publications which frequently provide *vitae* for author identification. In establishing the author names for its catalogs, Yale did so in relation not only to established forms of names on LC cards, but, perforce, in relation to its own catalog which included many more similar names. Inevitably the same surname has often been identified in either a briefer or fuller form, with or without dates, when one compares LC and Yale forms for the same individual. Neither can be said to be incorrect, yet they differ because they were established at different times to be compatible with different catalogs.

A card representing *Laws Relating to the Practice of Dentistry and Dental Hygiene* published by the Texas State Board of Dental Examiners and cataloged by the New York Public Library presents a number of variations from the cataloging data which the Library of Congress or many other libraries would provide. The NYPL main entry is "Texas. Statutes" while the LC form is "Texas. Law, Statutes, etc." Passing over variations from *Anglo-American Cataloging Rules* in capitalization and paragraphing in the body of the entry, one finds that the NYPL subject heading is "Dentistry—Jurisp.—U.S.—Texas" while the LC form is not only quite different but provides for direct rather than indirect subdivision. The added entry from NYPL is "Texas. Dental examiners, Board of" because document headings in its catalog have been established in an inverted form. The title is not classified but bears a unique, alpha-numeric number showing a fixed-order location. The difficulties in attempting to convert such entries to form and substance compatible with those of LC are obvious. The NYPL subject headings still retain in some substantial measure the early structure of an alphabetico-classed system. The Library of Congress uses "Malay Languages" as a subject heading with see-also references to some 55 related languages and dialects including "Tagalog." The NYPL form for "Tagalog" is "Malay language—Dialects: Tagala." These examples are not isolated exceptions in the entire body of cataloging contributed to the NUC in card form, but are representative of variations in a significant portion of the file.

A variety of classification schemes are represented in the NUC: LC, local adaptations of LC, Dewey, Dewey with changes and with numbers derived from many successive editions, and a host of other schemes. This latter category includes many locally developed or locally derived systems such as those of Yale, Harvard, NYPL, and many special libraries, as well as numbers for fixed-order systems and items with no classification at all. Many special libraries with significant holdings, such as Union Theological Seminary, the National Library of Medicine, and the National Agricultural Library, have their individual classification and subject heading systems. Together with major public and university libraries, including some of the largest contributors to the NUC card record, they have provided catalog records over the years which are often seriously incompatible with the data of other libraries and with LC cataloging. Again, it should be noted that the data are not incorrect but different.

Approximately 2.5 million catalog records (a gross sum, not adjusted for duplicates) have been converted to machine-readable form by 22 libraries.[14] Offhand, they seem to offer an inviting source of records for a national data base. But these libraries have encoded records that represent their individual cataloging experience and history. Although there may be a relatively high consistency within the data base of any one library or network, the records taken as a whole are unlikely to provide more than accidental consistency in terms of the entry forms, subject terminology, etc. They also represent differing levels of data, running from brief identification for purposes of automated circulation control to full bibliographic records compatible with MARC. Therefore, it is apparent that a major editing and recataloging effort would be required to assimilate them into a uniform data base.

The conclusion seems inescapable that the most useful machine-readable bibliographic data base must be one derived from a single major source, as is the current base being developed in the MARC program. It should be a source that offers a relatively high degree of consistency in the application of cataloging standards, one which reflects a full rather than a partial record, and one that has historically incorporated changes and is still hospitable to future change and updating. This confirms that the conversion of retrospective catalog records should, insofar as possible, be based on the LC Official Catalog record. Nevertheless, we need also to pursue solutions to the problem of how to expand and enhance the retrospective data base beyond the initial scope of the LC Official Catalog in order to incorporate the millions of titles not held by the Library of Congress. Cooperative funding, rather than cooperative preparation, may well be the route to follow.

References

[1] RECON Working Task Force. Conversion of retrospective catalog records to machine-readable form. Washington, Library of Congress, 1969. p. 13.

[2] Ibid., p. 11.

[3] Ibid., p. [138].

[4] Ibid., p. 107.

[5] Ibid., p. 20–34.

[6] Dawson, John M. "The acquisitions and cataloging of research libraries: a study of the possibilities for centralized processing." *Library Quarterly*, v. 27, January 1957, p. 11, 14, [20].

[7] Dewton, Johannes. "The grand illusion." *Library Journal*, v. 86, May 1, 1961, p. 1725.

[8] Library of Congress report to the General Education Board on the Cooperative Cataloging Project ending December 31, 1943. *In* American Library Association. Division of Cataloging and Classification. Catalogers' and classifiers' yearbook, no. 11. Chicago, 1945, p. 89.

[9] Dawson, *op. cit.*, p. 11.

[10] U.S. Library of Congress. Annual report of the Librarian of Congress for the fiscal year ending June 30, 1934. Washington, 1934. p. 194.

[11] U.S. Library of Congress. Annual report of the Librarian of Congress for the fiscal year ending June 30, 1941. Washington, 1942. p. 204.

[12] U.S. Library of Congress. Annual report of the Librarian of Congress for the fiscal year ending June 30, 1967. Washington, 1968. p. 136.

[13] Skipper, James E. "Future implications of Title IIC, Higher Education Act of 1965." *Library Resources and Technical Services*, v. 11, Winter 1967, p. 47.

[14] See Chapter 4, p. 7.

APPENDIX B

The National Union Catalog—Its Characteristics and Activity

This paper describes briefly the major characteristics of the National Union Catalog (NUC) maintained by the Library of Congress and gives basic data about the level of reporting by American libraries. This information may be helpful in analyzing some of the problems that must be faced in planning for a national bibliographic store in machine-readable form.

NUC as an entity is a file of catalog records for works held by American libraries. In general, each distinct record is represented by a single entry under author or title but added entry references are included in the newer part. Since NUC reports are subjected to only minimal editing, the same bibliographical item may be represented by more than one entry filed under different headings in widely dispersed portions of the file.

NUC is divided into two main components: the older part covers imprints through December 31, 1955; the new part covers 1956 and later imprints. When the NUC Publication Project began in 1967, it was estimated that the pre-1956 part contained 16–18 million cards. The proportion of duplicate entries was known to be high, however, and it has been confirmed in the process of editing and publishing volumes for the entries under A and B. It is probable that the true size of this part of NUC is closer to 10 million cards. The post-1956 file contains about 3.75 million cards (including references and added entries) for items not yet represented in a quinquennial book catalog.

Responsibility for reporting to NUC is assigned on a regional basis. An effort is made to have at least two libraries in each region report comprehensively; the others, selectively. Criteria for full reporting of 1956+ imprints and selective reporting are set forth in Addendum 1. The unit for reporting is "card" represented by LC printed cards, card order slips, or skeleton entries for items represented by LC cards. Titles not represented by LC printed cards are supposed to be reported in full cataloging form.

It is difficult to estimate the number of libraries represented in the NUC. The libraries listed in *Symbols of American Libraries* [1] are not a true indication of contributors because that publication provides symbols for many institutions that have not yet sent cards to NUC. A current estimate by the Chief of the Union Catalog Division places the number in the vicinity of 1,000. This figure takes as its base a 1962 statement that 763 libraries had reported their holdings up to that time.[2] The number of active contributors is much smaller, amounting to 328 libraries in fiscal 1969.[3] It should be

TABLE B.1—*Distribution of libraries reporting to the National Union Catalog, by number of reports and date of coverage, July 1, 1968 through June 30, 1969*

Number of reports	Number of libraries submitting reports of—		
	Any imprint date	Pre-1956 imprints	1956 and later imprints
Total	328	[1] 327	[1] 327
None		12	17
Less than 50	33	77	33
50 to 99	14	27	18
100 to 499	[2] 61	85	64
500 to 999	41	30	35
1,000 to 4,999	80	53	76
5,000 to 9,999	24	21	20
10,000 to 14,999	15	12	14
15,000 to 19,999	14	5	10
20,000 to 29,999	13	3	18
30,000 or more [3]	33	2	22

[1] Excludes the library mentioned in note 2.
[2] Includes 1 library that reported 315 titles for which no breakdown by imprint date is available.
[3] Apart from the number reported by the Library of Congress, the largest number of pre-1956 imprints was 53,851 and of 1956 and later imprints, 81,805; the largest total contribution was 135,656.
N. B. It is important to remember that, in this table, each of the columns represents a separate distribution so that the horizontal lines are not additive.

taken into consideration, however, that receipts from such sources as the Union Library Catalogue of the Philadelphia Metropolitan Area and the Cleveland Regional Union Catalog comprise titles from a number of libraries.

The level of reporting naturally varies considerably from library to library. Apart from the Library of Congress, the largest contributor reported nearly 136,000 titles and several contributors reported only one title. Table B.1 shows the distribution of active libraries by number of reports and date of coverage for fiscal 1969.

References

[1] U.S. Library of Congress. Union Catalog Division. Symbols of American libraries. 10th ed. Washington, 1969.

[2] U.S. Library of Congress. Annual report of the Librarian of Congress for the fiscal year ending June 30, 1961. Washington, 1962. p. 15.

[3] Data supplied by the Union Catalog Division.

Addendum 1

Criteria for Full Reporting of 1956 Imprints to The National Union Catalog Approved by the A.L.A. Board on Resources Committee on the National Union Catalog, Chicago, Jan. 30, 1957

To assure that the printed National Union Catalog will be developed to its full potentialities (i.e. to contain entries for all titles of 1956 imprints acquired by American libraries and to record approximately twenty locations of such works geographically dispersed throughout the U.S.A.) the A.L.A. Board on Resources has recommended that a relatively small number of important libraries in strategic geographical locations undertake "full" reporting and that hundreds of smaller, or special libraries provide "selective" reporting to *The National Union Catalog*. The following criteria for "full" reporting were approved by the Board on January 30, 1957.

The word "full" is not to be interpreted as "complete" or "entire", since there are certain categories for which cards would be superfluous. Thus, when a library is asked to report "fully" it should report all 1956 imprints, including those that are represented by LC printed cards, with the following exceptions:

Reprints
Serials
United Nations Publications
Titles for which "cdp" copy is requested by Card Division
Official state publications
(except the one library in each state designated to report)
U.S. Government Publications
(except analytics in series not analyzed on LC cards)

Of course, those libraries that duplicate *all* of their cards and find it expedient to send copies of all such cards may continue to do so—unnecessary cards will be discarded by the Union Catalog Division. However, if selection by the cooperating library will prove advantageous, cards for the above indicated categories of materials may be withheld with a resulting saving of labor at the National Union Catalog.

Be sure that the proper symbol for your library is affixed to each entry. Libraries that duplicate their own cards should add an asterisk to their library symbol when such cards are produced from unaltered LC card texts. This will expedite the handling of entries by the Editorial Staff. Cards should be sent to the Union Catalog Division, Library of Congress, Washington D.C. 20540. Yellow mailing labels are available on request.

The suggested categories of exclusion will be applicable to the general run of cataloged materials. However, it is expected that on occasion catalogers will recognize exceptional titles within these categories which should be reported because of their rarity, unusual research value, etc.

Items represented by LC printed cards may be reported in any of the following simplified forms:

Send yellow card order slips that are returned by the Card Division with filled card orders. These slips should be stamped "For NUC from ____".

Send a skeleton entry which may be limited to full author entry, first few words of title, imprint date, LC card number, and your library symbol.

Send a copy of the LC card on which the symbol for your library is affixed.

Revised Criteria for Selective Reporting of 1956 Imprints to the National Union Catalog Approved by the A.L.A. Board on Resources Committee on the National Union Catalog, Chicago, January 30, 1957

These criteria are devised to make certain that at least one copy of every title of potential research value published in 1956 and later is recorded in *The National Union Catalog* and at the same time

prevent a flood of reports of widely held commonplace books beyond the capacity of the editorial staff to handle. The immediate objective is to provide a published national catalog of monographs of 1956 imprints through *The National Union Catalog* and of serial publications, commencing with 1950, through *New Serial Titles*. Both publications attempt to locate titles in libraries at various geographical points throughout the U.S. and Canada so that the interlibrary loan burden will be spread more equitably and that borrowing libraries will have a reasonable chance of finding desired items in a neighboring institution.

The following are general criteria intended for the guidance of libraries asked to report only a selection of their 1956 imprints. Titles falling within the criteria are to be reported *even when LC printed cards are available.*

What To Report

Monographs (including monographs in series)

1. All books published *outside the U.S.* including titles in all alphabets and publications of foreign governments.

2. Items not in the book trade published in your region and/or within your sphere of acquisition.

3. Publications of the state government of the state in which your library is located unless another library in your state is reporting such materials (but not of other states).

4. In addition to the above broad categories, cards should be sent to the NUC for:

   a) All titles for which no LC cards are available.

   b) Imprints of rare or unusual character, or which are considered collectors' items.

   c) Analytics of monographs in series (including U.S. Government publications) when not analyzed by LC cards.

5. Revised entries for works previously reported should be clearly designated as such and should indicate previous form of entry when main heading has been changed.

Serials

New Serial Titles, the serials counterpart of *The National Union Catalog*, lists titles and holdings of serials whose first number was issued January 1, 1950 and later. Such entries will not be published in the NUC.

Borderline publications which might be cataloged as either monographs or serials may be reported to *The National Union Catalog* which will either publish the entry or forward it to *New Serial Titles.*

Libraries not now reporting to *NST* are urged to secure report forms and instructions from The Editor, *New Serial Titles*, Library of Congress, Washington 25, D.C.

Note: Catalog cards should be sent to the Union Catalog Division, Library of Congress, Washington 25, D.C. Yellow mailing labels are available on request.

How To Report

All reports to the NUC should be identified with the proper library symbol.

Titles not represented by LC printed cards should be reported in full cataloging form, including added and subject entries.

Items that are represented by LC printed cards may be reported in any of the following simplified forms:

Send yellow card order slips that have been returned to you by the Card Division with filled card orders. Such slips should be stamped "For NUC from ——". Or,

Send skeleton entries giving full main heading, first few words of title, imprint date, LC card number, and your library symbol. Or,

Send copies of LC printed cards on which your symbol is affixed.

Libraries that do not use LC cards are urged to add an asterisk to their library identification symbol when such cards are produced from unaltered LC card texts. This practice will expedite the handling of such entries by the Editorial Staff.

The National Union Catalog
General Information

The National Union Catalog (NUC) is a record of publications and their location in the Library of Congress and more than 1,100 other libraries in the United States and Canada. As such it is the central register of library resources in North

America. Major portions of the NUC are published on a continuing basis as detailed in paragraphs 3 and 4, but the bulk of the record for imprints prior to 1956 is contained in card files. This NUC on cards is housed principally in the Library's Main Building, Room MB–140–A. Until its abolition in July 1970 the Union Catalog Division exercised most NUC functions, including liaison with the public, but the various activities relating to the NUC are currently distributed among several Library divisions. The following statement summarizes present arrangements. Further information concerning any of the following services or publications is available upon request to the appropriate address.

1. Reference Service

Reference service on book locations and bibliographic information recorded in the NUC (published and unpublished) and in various auxiliary union catalogs in oriental and Slavic languages is the responsibility of the Union Catalog and International Organizations Reference Section, General Reference and Bibliography Division. The office of Robert W. Schaaf, Head of the Section, and John W. Kimball, Assistant Head, is located in MB–144 Balcony (phone 202–426–5534). The Union Catalog Reference Unit, Mrs. Dorothy Kearney, Supervisor, is in the adjacent Room MB–140–A which houses most of the NUC card files for imprints prior to 1952. As part of its service the Unit prepares and circulates to about 75 research libraries the *Weekly List of Unlocated Research Books*. The telephone number for reference inquiries is 202–426–6300. Written requests should be addressed: Library of Congress, Union Catalog Reference Unit, Washington, D.C. 20540.

2. Submission of Reports to NUC

Matters concerning reports to the NUC (i.e., the transmission of catalog cards of any imprint date by libraries to the NUC), and replies to inquiries concerning reporting criteria are the responsibility of the Catalog Publication Division, Mrs. Gloria Hsia, Chief. This division is located in the Massachusetts Avenue Annex, 214 Massachusetts Avenue NE. The address is Library of Congress, Catalog Publication Division, Washington, D.C. 20540.

3. Catalog of Post-1955 Imprints and Specialized Publications

The National Union Catalog, a Cumulative Author List is published in monthly issues with quarterly, annual, and quinquennial cumulations. It includes titles currently cataloged by the Library of Congress on printed cards and monographic titles for 1956 and later years that are reported by major U.S. research libraries and some Canadian libraries. This *Catalog* is supplemented by the *Register of Additional Locations*. Other specialized publications are:

Symbols of American Libraries (Earlier editions entitled: *Symbols Used in the National Union Catalog of the Library of Congress*)

Requests for symbols for additional libraries to be included and notices of changes of name, etc., of libraries already included, should be addressed to Library of Congress, Catalog Publication Division, Editor, Symbols of American Libraries, Washington, D.C. 20540.

National Register of Microform Masters

Reports of locations of microform masters (i.e., microforms used only to make other copies) should be addressed to Library of Congress, Catalog Publication Division, Editor, National Register of Microform Masters, Washington, D.C. 20540.

Newspapers on Microfilm

Reports of microfilms of American and foreign newspapers should be addressed to Library of Congress, Catalog Publication Division, Editor, Newspapers on Microfilm, Washington, D.C. 20540.

Microfilming Clearing House Bulletin

This is issued at irregular intervals, as reports are received, and appears as a supplement to the *Library of Congress Information Bulletin*. Reports of major microfilming projects, planned or completed, should be made to Library of Congress, Catalog Publication Division, Microfilming Clearing House, Washington, D.C. 20540. In addition to general reports of projects to MCH, reports of the individual titles filmed as part of the project should also be made to the editor of the pertinent Library of Congress catalog.

The *National Register of Microform Masters* is edited by Harold Cumbo (202-426-5980). *Newspapers on Microfilm*, *the Microfilming Clearing House Bulletin*, and *Symbols of American Libraries* are edited by Imre Jarmy (202-426-5959).

4. National Union Catalog, Pre-1956 Imprints

The National Union Catalog Publication Project, Johannes Dewton, Head, is responsible for editing the *National Union Catalog, Pre-1956 Imprints*, which is being published by Mansell Information/Publishing Ltd. Over 100 of a projected 610 volumes have been issued, and the entire project is expected to take about 10 years. Staff of the project, which is not charged with responsibility for service to the public, is located in MB-137.

479-312 O—73——4

APPENDIX C

Major Duties Involved in the Preparation of the Library of Congress Book Catalogs

The following list summarizes the major duties involved in the manual preparation of the Library of Congress book catalogs:

A. Arranging and sorting

For National Union Catalog and Register of Additional Locations

1. Receiving, recording, and sorting of outside library reports.
2. Receiving, recording, and sorting of LC printed cards.
3. Recording and sorting of typed printing file cards.
4. Sorting of various other cards such as cancellation notices, entry revision notices, etc.
5. Arranging all of the above, some numerically, for processing or filing.

For Books: Subjects

1. Sorting of LC printed cards, typed subject heading and reference cards, also cards for various in-process or auxiliary files.
2. Arranging these for processing or filing.

For other catalogs

Each catalog has its own array of print files, auxiliary files and in-process files for which cards may be sorted, recorded, or arranged.

B. Filing

For NUC

1. Filing into Control File.
2. Filing into several print files.
3. Filing into various in-process or auxiliary files.

For Books: Subjects

1. Filing into subject authority file.
2. Filing into several print files.
3. Filing into various in-process or auxiliary files.

For other catalogs

1. Filing into several print files.
2. Filing into various in-process and auxiliary files.

C. Searching

For NUC and Register

1. Searching in Control File and 1958–1962 printed book catalog.
2. If found, add symbol to Control File card and forward report; current to NUC *Author List*, non-current to *Register*.
3. If not found, but heading is found, refer to be edited.
4. If not found and heading lacking, either a) if modern personal author heading, refer to be edited; or b) if corporate author or older personal name heading, refer to Official Catalog for additional searching.
5. Searching in Official Catalog, when required, for established form of heading for corporate authors and older personal names.
6. Searching and matching in NUC print files to add current locations.
7. For the Register, searching in the NUC 1963 annual to add card number to 1963 outside library reports.
8. Searching conflicts in Official Catalog, the Control File, the book catalogs, or various Card Division files.

For NRMM

Searching for LC card numbers and established headings in Official Catalog, NUC: *Pre-1956 Imprints*, the Main Catalog, etc.

For other catalogs

Various searching to determine status of particular entry or heading, to solve conflicts, to establish headings.

D. Editing

For NUC (Outside Library Reports)

1. Verifying choice and form of heading. Establish heading if new.
2. Verifying general correctness of cataloging.
3. Providing for requisite added entries and cross references.

For NUC (Printing files and/or Control File)

1. Providing for requisite added entries and cross references for LC cards.
2. Providing for requisite information cards: history cards, name prefix cards, acronym cards, etc.
3. Reviewing of print files and final page copy.
4. Solving conflicts, correcting errors, updating entries, making corrections and changes. Coordinating related affected entries in same file. Coordinating changes between the printing files and the Control File and between Catalog Publication Division and the descriptive cataloging divisions.

For Register

1. Preparing brief author-title entries for added locations to outside library reports in the NUC 1958–1962 issue.
2. Preparing controls and references for cancelled or superseded card numbers.
3. Reviewing of print files and final page copy.
4. Solving conflicts, correcting errors, updating entries, making corrections and changes. Coordinating changes between the *Register* files, the NUC print files, and the Control File.

For other catalogs

Each catalog has its own editing requirements based on catalog content, entry format, etc.

E. Typing and proofreading

For NUC

1. Typing of printing file cards for outside library reports.
2. Typing of printing file added entries and cross references for LC printed main entries.
3. Proofreading of typed cards. (Typed added entries and cross references for outside reports are also Xeroxed for use in Control File).

For other catalogs

As required.

F. Composing of page copy (Mounting and stripping)

For all catalogs (except *Symbols of American Libraries*)

1. Preparing camera copy by shingling and taping cards onto 14½ by 20 inch cardboards.
2. Numbering and collating pages.
3. Dismantling of camera copy used in past issues so that cards can be re-used in the next larger cumulation.

APPENDIX D

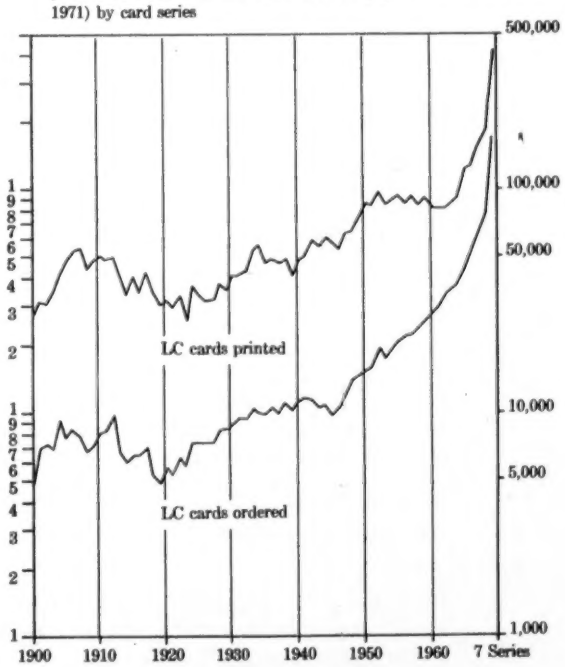
Analysis of Library of Congress Card Orders (April 1970-March 1971)

The Card Division provided a magnetic tape listing all LC titles ordered in a one-year period and the frequency of order. From April 1, 1970, to March 31, 1971, 11,896,521 orders were received for 1,209,198 titles. Tables were made to summarize the entire tape and two subsets of the tape. The first group was a random sample of 1,710 titles selected to study the relationship of language to card orders. The second group comprised the 1,000 most frequently ordered cards. It should be remembered that this analysis does not include the over 1,000 subscribers to complete sets of LC proofsheets or the 84 research libraries who receive depository sets of all currently printed LC cards. If these libraries had ordered cards instead, the number of cards ordered from the 7 series would probably be substantially larger.

The following comments point out any unusual characteristics of each table. Because some of the counts were made by computer and some by hand or estimation, the tables have certain small discrepancies. Figures have been rounded to indicate the approximations. Because the 7 series began in December 1968, the 7 series includes both 1969 and 1970 printed cards. It was decided to ignore the number of cards printed between January and March 1971 as being too recent to have been ordered by outside libraries.

Tables D.1–3 and Figure D.1 present some characteristics of the total tape. Table D.1 shows that 42 percent of the 7-series cards printed were ordered. Actually the demand for current LC cards was appreciably higher as reflected by the distribution of proofsheets and depository sets. The number of cards ordered once (as shown in Tables D.2 and D.3) differs by 0.25 percent. The almost constant level of cards ordered one time is shown in Table D.3. Figure D.1 shows the close relationship of cards printed (top line) to cards ordered (bottom line).

FIGURE D.1—Number of LC cards printed by card series in comparison to number of LC cards ordered (April 1970–March 1971) by card series



Tables D.4–6 are based on a random sample of 1,710 cards which was drawn from the entire listing of LC card orders. The fact that the total percentages in Tables D.5 and D.6 are remarkably similar to those in Tables D.1 and D.2 confirms that the sample is representative of the total. A rough count showed that less than two percent of the cards were not monographs; almost all the serials were in English. Because this analysis covers only card orders and does not include the use of proofsheets, depository sets or book catalogs, the high percentage of English titles ordered (77

TABLE D.1—*Number and percentage of all titles ordered from April 1970 through March 1971, by period of series*

Period of series	Number of titles ordered [1]	Percentage of total titles ordered	Total number of titles available [1]	Percentage ordered	Average number of orders [2]
Pre-1900	3,400	0.3	20,500	16.6	2.5
1900–09	69,700	5.8	394,400	17.7	2.2
1910–19	69,000	5.7	406,200	17.0	2.2
1920–29	68,700	5.7	329,900	20.8	2.7
1930–39	99,500	8.2	468,200	21.3	3.0
1940–49	116,100	9.6	592,700	19.6	3.6
1950–59	203,500	16.8	896,200	22.7	5.7
1960–68	405,200	33.5	1,029,100	39.4	11.5
7 series [3]	174,100	14.4	416,900	41.8	27.9
All series	1,209,200	100.0	4,554,100	26.6	9.8

[1] Data rounded to nearest hundred.
[2] Calculated from unrounded data.
[3] Includes 1969 and 1970 cards.

TABLE D.2—*Number and percentage of titles ordered between April 1970 through March 1971, by frequency of orders*

Frequency of orders	Number of titles [1]	Percentage	Cumulative	
			Number of titles	Percentage
2,000 or more	[2] 20	(3)	20	(3)
1,000 to 1,999	110	(3)	130	0.01
900 to 999	40	(3)	170	.01
800 to 899	60	(3)	230	.02
700 to 799	100	0.01	330	.03
600 to 699	180	.01	510	.04
500 to 599	310	.03	820	.07
400 to 499	620	.05	1,440	.12
300 to 399	1,400	.12	2,840	.23
200 to 299	3,500	.29	6,340	.52
100 to 199	12,800	1.06	19,140	1.58
90 to 99	3,000	.25	22,140	1.83
80 to 89	4,200	.35	26,340	2.18
70 to 79	5,000	.41	31,340	2.59
60 to 69	7,300	.60	38,640	3.20
50 to 59	10,100	.84	48,740	4.03
40 to 49	13,800	1.14	62,540	5.17
30 to 39	22,300	1.84	84,840	7.02
20 to 29	38,700	3.20	123,540	10.22
10 to 19	92,900	7.68	216,440	17.90
9	18,500	1.53	234,940	19.43
8	22,500	1.86	257,440	21.29
7	27,100	2.24	284,540	23.53
6	35,600	2.94	320,140	26.47
5	46,000	3.80	366,140	30.28
4	62,500	5.17	428,640	35.45
3	99,100	8.20	527,740	43.64
2	195,100	16.13	722,840	59.78
1	486,400	40.22	1,209,240	100.00

[1] Figures above 1,000 rounded to tens; those below 1,000 rounded to hundreds.
[2] The largest number of orders for a title was 3,280.
[3] Less than 0.01 percent.

TABLE D.3—*Number and percentage of titles ordered once from April 1970 through March 1971, by period of series*

Period of series	Number of titles ordered [1]	Percentage of total titles ordered	Total number of titles available [1]	Percentage ordered
Pre-1900	2,000	0.4	20,500	9.8
1900–09	41,100	8.4	394,400	10.4
1910–19	41,000	8.4	406,200	10.1
1920–29	36,800	7.6	329,900	11.2
1930–39	52,000	10.7	468,200	11.1
1940–49	58,700	12.0	592,700	9.9
1950–59	82,400	16.9	896,200	9.2
1960–68	124,400	25.5	1,029,100	12.1
7 series [2]	49,200	10.1	416,900	11.8
All series	487,600	100.0	4,554,100	10.7

[1] Data rounded to nearest hundred.
[2] Includes 1969 and 1970 cards.

TABLE D.4—*Number and percentage of cards in a sample of LC card orders, by language*

Language(s)	Number	Percentage	Percentage of current LC cataloging
English	1,320	77	37
French/German	170	10	17
Italian/Spanish/Portuguese/ Romanian	92	5	10
Dutch/Scandinavian	13	1	5
Russian	55	3	12
Other roman	28	2	5
Other nonroman	32	2	14
Total	1,710	100	100

TABLE D. 5—*Number and percentage of English and non-English cards in a sample of LC card orders, by period of series*

Period of series	English titles		Non-English titles		Total	
	Number	Percentage	Number	Percentage	Number	Percentage
Pre-1900	3	100. 0			3	0. 2
1900–09	83	85. 6	14	14. 4	97	5. 7
1910–19	87	85. 3	15	14. 7	102	6. 0
1920–29	87	85. 3	15	14. 7	102	6. 0
1930–39	113	77. 4	33	22. 6	146	8. 5
1940–49	122	72. 6	46	27. 4	168	9. 8
1950–59	229	76. 9	69	23. 1	298	17 4
1960–68	427	77. 2	126	22. 8	553	32 3
7-series [1]	169	70. 1	72	29. 9	241	14. 1
Total	1, 320	77. 2	390	22. 8	1, 710	100. 0

[1] Includes 1969 and 1970 cards.

TABLE D. 6—*Number and percentage of English and non-English cards in a sample of LC card orders, by frequency of orders*

Frequency of orders	English titles		Non-English titles		Total	
	Number	Percentage	Number	Percentage	Number	Percentage
400 to 499	1	100. 0			1	0. 1
300 to 399	2	100. 0			2	. 1
200 to 299	5	100. 0			5	. 3
100 to 199	19	100. 0			19	1. 1
90 to 99	4	100. 0			4	. 2
80 to 89	6	100. 0			6	. 4
70 to 79	7	100. 0			7	. 4
60 to 69	11	100. 0			11	. 6
50 to 59	15	100. 0			15	. 9
40 to 49	21	100. 0			21	1. 2
30 to 39	29	93. 6	2	6. 4	31	1. 8
20 to 29	52	100. 0			52	3. 0
10 to 19	121	96. 8	4	3. 2	125	7. 3
9	23	85. 2	4	14. 8	27	1. 6
8	28	90. 3	3	9. 7	31	1. 8
7	35	89. 7	4	10. 3	39	2. 3
6	43	86. 0	7	14. 0	50	2. 9
5	62	89. 9	7	10. 1	69	4. 0
4	71	81. 6	16	18. 4	87	5. 1
3	115	81. 0	27	19. 0	142	8. 3
2	191	69. 7	83	30. 3	274	16. 0
1	459	66. 3	233	33. 7	692	40. 5
Total	1, 320	77. 2	390	22. 8	1, 710	99. 9

percent) does not mean that foreign language titles are not needed by American libraries.

Tables D.7 and D.8 are related to the 1,000 most frequently ordered cards. Eight of the printed cards were not available from the Card Division and two had been superseded by later revisions of the same titles. Therefore the 1,000 most frequently ordered cards were reduced to 990 titles as shown in the tables. All 990 cards were English and 90 percent (887) had the MARC notation on them. The range of orders was from 3,280 to 470. Ninety-four percent (933) were monographs; 6 percent (55) were serials and two titles were atlases.

TABLE D. 7—*Number and percentage of 1,000 most frequently ordered cards, by period of series*

Period of series	Number	Percentage
1904–09	5	0. 5
1910–19	2	. 2
1920–29	3	. 3
1930–39	4	. 4
1940–49	5	. 5
1950–59	7	. 7
1960–68	91	9. 2
7-series [1]	873	88. 2
Total	990	100. 0

[1] Includes 1969 and 1970 cards.

TABLE D. 8—*Number and percentage of 1,000 most frequently ordered cards, by LC classification*

LC classification	Number	Percentage	Percentage of current LC cataloging
A	5	0. 5	0. 8
B	42	4. 2	7. 5
C	14	1. 4	1. 1
D	61	6. 2	10. 7
E	136	13. 7	1. 6
F	14	1. 4	2. 1
G	22	2. 2	2. 9
H	187	18. 9	13. 3
J	17	1. 7	2. 5
KF	15	1. 5	3. 1
L	62	6. 3	3. 3
M	19	1. 9	2. 8
N	21	2. 1	4. 5
P	179	18. 1	22. 5
Q	68	6. 9	6. 1
R	29	2. 9	2. 1
S	12	1. 2	1. 6
T	38	3. 8	8. 5
U	14	1. 4	. 4
V	2	. 2	. 3
Z	32	3. 2	2. 2
Law	1	. 1	([1])
Total	990	99. 8	99. 9

[1] Figure not available.

Index

Agency for RECON, 3, 32

Bibliographic problems, 33–36
Book catalog procedures, 42–43

Catalog comparison, 15
Cataloging, variations in, 34–35
Content designators, 4–5
Conversion of other machine-readable data bases: assumptions about, 12–13; costs, 2, 11–12, 14–17; need for editing, 12, 36; objectives, 12; potential savings from, 11; procedures, 14–15; programming, 11; strategy, 12–14; system considerations, 16
Conversion strategy, 3, 30–32
Cooperative cataloging, 34–35

Distribution function in relation to level of MARC record, 5

Elimination of ineligible records from other data bases, 13–15

Format recognition, 11, 14–15
Funding, 3, 32

Inconsistencies in NUC reports, 34–35
Indexes as an alternative to full conversion, 31
Indexes to NUC: cumulation, 22; data elements, 20–21; master index record, 21–22; sorting, 24–25; types, 19–20; updating, 23

Keying; see Typing

LC card orders, analysis of, 44–47
LC catalog cards, changes by other libraries, 34
Levels of machine-readable records, 2, 4–6, 31; definition, 4
Location reports for NUC, 19, 23–24, 38–41

Manpower costs, 15
MARC Distribution Service, 22, 23
MARC records in other data bases, 7
Master index record, 21–22
Microstorage with machine indexes, 31
Model NUC network, 28

National Commission for Libraries and Information Science, 3, 32
National Union Catalog (NUC), automated: advantages, 2, 29; components, 19, 22; cost factors, 26–28; design, 19–20; location reports, 23–24; machine files, 24–25; publication pattern, 22–23; register, 19–21; system, 22–26; types of indexes, 19–20; updating, 24–25
National Union Catalog, manual: Author List, 19; bibliographic problems, 18, 34–35; Books: Subjects, 19; characteristics, 18, 33–34, 37–38; components, 19, 40–41; criteria for evaluation of, 18; criteria for reporting to, 38–40; effect of automation, 26–27; libraries reporting to, 37–38; Register of Additional Locations, 19
Network of regional centers, 28
New York Public Library, 15
Nonsystematic conversion, 30–31
NUC function in relation to level of MARC record, 6

Other machine-readable data bases: characteristics, 10, 15; machine formats, 8–9; potential, 9; recommendations about, 32; standard for reporting on, 2, 16–17; survey, 7–11

Programming costs, 11
Proofing, 13, 15

RECON Advisory Committee, v
RECON Pilot Project, 30
RECON studies: funding, iii; goals, 1
RECON study, original, 1, 4, 33
RECON Working Task Force, v
Register numbers, 21

Sorting, 24–25
Systematic conversion, 30–31

Typing, 13, 15

Unit costs; see Manpower costs
University of Chicago Library, 14–15
Updating, 15, 23–25

Verifying, 13, 15

U.S. GOVERNMENT PRINTING OFFICE: 1973 O—479-312